



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2009

---

## **Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse**

Bubenhofer, Noah

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-111287>

Monograph

Published Version

Originally published at:

Bubenhofer, Noah (2009). Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse. Berlin, New York: de Gruyter.

Kopie im Originallayout von:

Bubenhofer, Noah (2009): Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse. Berlin/New York: De Gruyter (=Sprache und Wissen, 4).

NOAH BUBENHOFER

# Sprache<sup>und</sup> Wissen

## Sprachgebrauchsmuster

Korpuslinguistik als Methode  
der Diskurs- und Kulturanalyse

DE  GRUYTER



# **Sprachgebrauchsmuster**

**Korpuslinguistik als Methode der Diskurs- und  
Kulturanalyse**

Noah Bubenhofer

30. April 2009





IV

Seite 4 (Verlag)

Bitte einfügen:

*Die vorliegende Arbeit wurde von der Philosophischen Fakultät der Universität Zürich im Herbstsemester 2008 auf Antrag von Prof. Dr. Angelika Linke und Prof. Dr. Martin Volk als Dissertation angenommen.*

## Vorwort

Dieses Buch gäbe es nicht, ohne eine Reihe von Personen, die mich unterstützt haben. Allen voran Prof. Dr. Angelika Linke und Prof. Dr. Martin Volk. Sie haben mich motiviert, diesen Weg einzuschlagen und begleiteten mich fachlich, aber auch menschlich. Angelika Linke sorgte darüber hinaus für eine ideale Arbeitsumgebung und Martin Volk begeisterte mich für die Computer- und Korpuslinguistik. Und beide ließen sich auf das Wagnis einer Arbeit zwischen den Disziplinen ein; ihre Offenheit dem jeweils Fachfremden gegenüber machte dies erst möglich.

Ein solches Projekt kann aber nur in einem Umfeld von Freunden, Kolleginnen und Kollegen gedeihen, wie ich es am Deutschen Seminar der Universität Zürich vorfand. Dazu gehören Joachim Scharloth, Jürgen Spitzmüller, Martin Businger, Sascha Demarmels, Juliane Schröter, Daniela Wagner-Macher, Jacqueline Holzer, Wolfgang Kesselheim, Britta Juska-Bacher, Thomas Forrer, Vanessa Simili, Markus Domeisen, Ursula Landert und einige mehr. Sie diskutierten mit mir fachliche Ideen und Probleme, gaben Ratschläge und waren wichtig, weil sie ähnliche Situationen bereits hinter sich hatten, auch gerade mittendrin waren oder es aber gerade nicht waren.

Methodisch besonders inspirierend waren die Arbeiten der korpuslinguistisch arbeitenden Kolleginnen und Kollegen am Institut für Deutsche Sprache in Mannheim. Kathrin Steyer, Cyril Belica, Holger Keibel, Marc Kupietz, Rainer Perkuhn und Annelen Brunner verhalfen mir zu wichtigen Einsichten, gaben mir Einblick in ihre Forschungsarbeiten und sind deshalb zu wichtigen Diskussionspartnern geworden.

Dominik Heeb verhalf mir zu statistischer Bodenhaftung und wachte darüber, dass ich die Regeln der Kunst nicht zu arg strapazierte. Marcel Eggler bewahrte immer eine gesunde Skepsis gegenüber korpuslinguistischen Methoden; gerade deshalb waren die Diskussionen mit ihm besonders hilfreich. Stefaniya Ptashnyk half mir in der Endphase des Projekts, den Schlusspunkt zu finden. Dank geht an die



„Neue Zürcher Zeitung“ (NZZ Medien Verlag), die mir die Nutzung der Daten für die vorliegende Arbeit erlaubt hat.

Prof. Dr. Ekkehard Felder, Prof. Dr. Markus Hundt und Marcus Müller machten die Publikation bei Walter de Gruyter möglich. Felicitas Jungi ist für die umsichtige Korrektur der Arbeit verantwortlich. Beim Verlag sorgten Heiko Hartmann und Manuela Gerlof für eine exzellente Betreuung.

Weit zurück reicht die Freundschaft zu meinen Basler Kommilitonen Stefan Bertschi und Till A. Heilmann. Obwohl fachlich in unterschiedlichen Booten unterwegs, befanden wir uns auf der gleichen Passage und konnten uns so gegenseitig unterstützen. Während des Studiums in Basel begeisterte mich Prof. Dr. Heinrich Löffler für die Linguistik und ermutigte mich damals, dem Fach nach dem Studium nicht den Rücken zu kehren.

Ganz besonders wichtig sind aber meine Familie, meine Eltern, Heidi und Alex. Sie haben mich großzügig bei allen Plänen unterstützt und schon am Familientisch für eine Diskussionskultur gesorgt, die wissenschaftliches Arbeiten überhaupt erst ermöglicht.

Trotz allen diesen Menschen: Dieses Buch gäbe es nicht, wenn Ruth Schildknecht nicht immer an meiner Seite gestanden wäre.

All diesen Menschen sei an dieser Stelle herzlich gedankt.

*Zürich, April 2009*

## Inhalt

1	Einleitung . . . . .	1
1.1	Eingrenzung . . . . .	4
1.2	Hypothesenbildung . . . . .	5
1.3	Terminologie und allgemeine Hinweise . . . . .	8
<b>I</b>	<b>Muster als Effekte des Sprechens: Theoretische Perspektive</b>	<b>11</b>
2	Grundbegriffe . . . . .	15
2.1	Sprachgebrauch korpuslinguistisch . . . . .	15
2.2	Muster . . . . .	18
2.2.1	Bedeutungsvielfalt . . . . .	18
2.2.2	Klärung: Beispiele und Definition . . . . .	21
2.2.3	Muster als Muster erkennen . . . . .	24
2.2.4	Abgrenzungen und Fazit . . . . .	29
2.3	Diskurse . . . . .	31
2.3.1	Foucault'sche Diskursanalyse . . . . .	31
2.3.2	Linguistische Diskursanalyse . . . . .	32
2.3.3	Kritik . . . . .	35
2.3.4	Fazit . . . . .	38
2.4	Kultur . . . . .	39
2.4.1	Kultur als Bedeutungsgewebe . . . . .	39
2.4.2	Kultur als semiotisches System . . . . .	40
2.4.3	Fazit . . . . .	42
3	Die Bedeutung musterhaften Sprachgebrauchs . . . . .	43
3.1	Sprachgebrauchsmuster als Effekte sozialen Sprach- handelns . . . . .	43
3.1.1	Formelhafte Textoberfläche – Kontextualisierung	43
3.1.2	Sozialität und Kulturalität von Sprache . . . . .	46

3.1.3	Common Sense . . . . .	50
3.2	Sprache lesen, Kultur lesen . . . . .	52
4	Was sind Sprechweisen? . . . . .	55
4.1	Stil . . . . .	55
4.2	Kommunikative Gattungen . . . . .	57
4.3	Mentalitäten und diskurssemantische Grundfiguren . . . . .	60
4.4	Typik des Verhaltens und idiomatische Prägungen . . . . .	62
4.5	Argumentationsfiguren, Topoi . . . . .	67
4.6	Metaphern . . . . .	77
4.7	Sprachgebrauchsgeschichte als Kulturanalyse . . . . .	84
4.8	Die Einordnung der Konzepte . . . . .	87
<b>II</b>	<b>Mustererkennung als Basis für die Analyse:</b>	
	<b>Methodische Herleitung</b>	<b>95</b>
5	Grundsätze . . . . .	99
5.1	Daten lesen: ‚corpus-based‘ und ‚corpus-driven‘ . . . . .	99
5.2	Daten, Muster, Beschreibung: Die Heuristik im Überblick . . . . .	102
5.3	Wörter, Sätze, Texte? Die Elemente und Grenzen eines Diskurses bestimmen . . . . .	105
6	Musterhafte Strukturen finden: Die Klumpen im Text . . . . .	111
6.1	Kollokationen, Kookkurrenzen, n-Gramme . . . . .	111
6.2	Syntagmatische Muster . . . . .	118
6.3	Zwischenfazit: Terminologische Festlegung . . . . .	121
6.4	Kollokationen, corpus-driven . . . . .	123
6.5	Lemmata und Wortarten: Gewinn oder Verlust an Information? . . . . .	124
7	Der statistische Zugang zu sprachlichen Daten . . . . .	131
7.1	Deskriptive Statistik . . . . .	132
7.2	Testen von Hypothesen . . . . .	134
7.2.1	Klumpen: Wörter sind nie zufällig verteilt . . . . .	134
7.2.2	Kontingenztafel erstellen . . . . .	136
7.2.3	Chi-Quadrat-Test . . . . .	137

7.2.4	Log-Likelihood $G^2$ . . . . .	139
7.2.5	Weitere Tests . . . . .	139
7.3	Weitere Hilfsmittel . . . . .	142
7.3.1	Dispersion: Juillands D . . . . .	142
7.3.2	Differenzkoeffizient . . . . .	144
7.4	Fazit . . . . .	144
8	Von corpus-driven zu corpus-based . . . . .	149
8.1	Typikprofile erstellen . . . . .	149
8.2	Interpretieren: Diskurse modellieren . . . . .	152
8.2.1	Grad der Typik . . . . .	155
8.2.2	Kontextualisierungsprofile . . . . .	160
8.2.3	Paradigmatische Konstellationen . . . . .	163
8.2.4	Pragmatische Funktionen . . . . .	168
8.3	Überprüfen: Zurück in die Daten . . . . .	171
9	Ressourcen der Korpus- und Computerlinguistik . . . . .	175
9.1	Öffentlich verfügbare Referenzkorpora . . . . .	175
9.2	Software zur Korpusabfrage . . . . .	180
9.3	Software für die Berechnung von Kollokationen . . . . .	181
<b>III</b>	<b>Anwendungsbeispiele</b> . . . . .	<b>185</b>
10	Die Datengrundlage: Das NZZ-Korpus . . . . .	189
10.1	Publizistische Daten zur ‚Neuen Zürcher Zeitung‘ . . . . .	189
10.2	Eckdaten des Korpus . . . . .	190
11	Aufbereitung der Daten . . . . .	193
12	Berechnung der typischen Mehrworteinheiten . . . . .	197
12.1	Festlegung von Parametern und Teilkorpora . . . . .	197
12.2	Selektion der Mehrworteinheiten für die Analyse . . . . .	200
13	Beispielanalysen . . . . .	209
13.1	Die Auslandsberichterstattung im NZZ-Korpus . . . . .	209
13.1.1	Grobanalyse . . . . .	209
13.1.2	Die <i>bosnischen Serben</i> : Die Verwendung von Ethnienbezeichnungen . . . . .	212

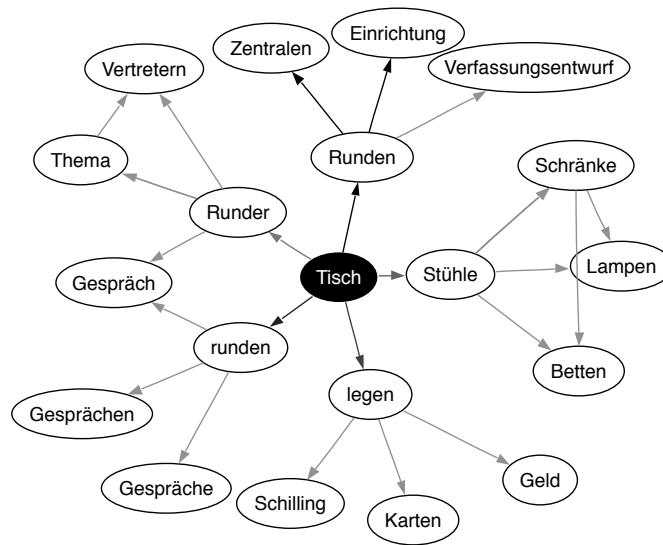
13.1.3	Kontextualisierungsprofile: <i>Kampf gegen X, Kampf dem X oder Kampf mit X?</i> . . . . .	220
13.1.4	Was ist wert, gezählt zu werden? <i>Die Zahl der X</i> . . . . .	229
13.2	Die Inlandsberichterstattung im NZZ-Korpus . . . . .	240
13.2.1	Grobanalyse . . . . .	241
13.2.2	Routinisierung der medialen Selbstreflexion: <i>vor den Medien</i> . . . . .	242
13.2.3	Das Sorgenbarometer auf der Textoberfläche: <i>Kampf gegen X, Bekämpfung von X</i> . . . . .	248
13.2.4	Argumentieren und bewerten: <i>nicht nur</i> <i>... sondern auch</i> . . . . .	257
13.3	Weitere Ressorts des NZZ-Korpus . . . . .	268
13.3.1	<i>Zum ersten Mal und die moderne, westliche, bürgerliche Gesellschaft</i> . . . . .	269
13.3.2	<i>Es ist nicht mehr wie früher.</i> . . . . .	269
13.3.3	Blick in die Vergangenheit: <i>die -iger Jahre</i> . . . . .	272
13.3.4	Veränderungen in der Wirtschaftswelt . . . . .	273
13.3.5	Geografische Referenzen . . . . .	274
13.3.6	Der nördliche Nachbar: <i>Die deutsche(n) X</i> . . . . .	280
13.3.7	Vom Schweizer Skiverband zu <i>Swiss Ski</i> . . . . .	285
13.3.8	Die <i>Damen</i> und die <i>Frauen</i> im Sport . . . . .	287
13.3.9	Ressorts im synchronen Vergleich: Sprachgebrauch in Leserbriefen . . . . .	291
14	Hypothesenbildung . . . . .	295
14.1	Sprechen über KRIEG UND GEWALT . . . . .	296
14.2	Sprechen über SORGEN UND PROBLEME . . . . .	298
14.3	Sprachgebrauch und Welt . . . . .	300
14.4	Sprachgebrauch und Text . . . . .	302
IV	Fazit und Ausblick . . . . .	305
15	Theorie und Praxis . . . . .	309
15.1	Sprachgebrauchsmuster als Kristallisationkerne von Diskursen . . . . .	309
15.2	Semantische Konzepte in der korpuslinguistischen Operationalisierung . . . . .	312

16	Praxis und Theorie . . . . .	317
16.1	Der Gewinn des korpuslinguistischen Blicks für die Theoriebildung . . . . .	317
16.1.1	Diskurse, Sprechweise, Themen . . . . .	317
16.1.2	Kontingente Diskursdefinitionen . . . . .	318
16.1.3	Semantische Matrizen typischer Sprechweisen . . . . .	319
16.2	Nutzen einer corpus-driven-Analyse für die Diskurs- und Kulturanalyse . . . . .	320
16.2.1	Korpuslinguistik und Diskursanalyse . . . . .	320
16.2.2	Korpuslinguistik und Kulturanalyse . . . . .	322
17	Ausblick . . . . .	325
17.1	Andere Daten, ergänzte Methodik: Ein Beispiel . . . . .	325
17.2	Schlüsse, Möglichkeiten, Desiderata . . . . .	332
17.2.1	Korpora . . . . .	332
17.2.2	Statistik . . . . .	333
17.2.3	Semantische Kategorisierungen und Annotation . . . . .	335
17.2.4	Sprachgebrauchsmuster als kleinste semantische Einheiten . . . . .	336
<b>V</b>	<b>Anhang</b>	<b>339</b>
A	Typikprofile . . . . .	341
A.1	Ressort ‚Ausland‘ . . . . .	341
A.1.1	Periode 1995–1997 . . . . .	341
A.1.2	Periode 2003–2005 . . . . .	343
A.2	Ressort ‚Inland‘ . . . . .	345
A.2.1	Periode 1995–1997 . . . . .	345
A.2.2	Periode 2003–2005 . . . . .	347
A.3	Ressort ‚Feuilleton‘ . . . . .	350
A.3.1	Periode 1995–1997 . . . . .	350
A.3.2	Periode 2003–2005 . . . . .	351
A.4	Ressort ‚Wirtschaft‘ . . . . .	352
A.4.1	Periode 1995–1997 . . . . .	352
A.4.2	Periode 2003–2005 . . . . .	355
A.5	Ressort ‚Sport‘ . . . . .	357
A.5.1	Periode 1995–1997 . . . . .	357

A.5.2 Periode 2003–2005 . . . . .	358
A.6 Ressort ‚Leserbriefe‘ . . . . .	359
Tabellenverzeichnis . . . . .	361
Abbildungsverzeichnis . . . . .	365
Literatur . . . . .	369
Nachschlagewerke . . . . .	381
Quellentexte und Korpora . . . . .	383
Stichwortverzeichnis . . . . .	385

## 1 Einleitung

Der Ausgangspunkt für die vorliegende Untersuchung ist eine Grafik, die typischen Sprachgebrauch visualisiert:



**Abbildung 1.1:** Kollokatoren zu *Tisch* im DeReKo IDS (o.J.)-Korpus. Auszug aus der Kookkurrenzdatenbank CCDB (Belica 2001–2006); Visualisierung NB.

Das Wort *rund* tritt überzufällig häufig mit dem Wort *Tisch* zusammen auf. Diese Feststellung gilt für ein zwei Milliarden Wörter umfassendes Korpus: Das DeReKo IDS (o.J.)-Korpus. In Abbildung 1.1 sind weitere wichtige Verbindungen von *Tisch* zu anderen Wörtern dargestellt. Die Grafik zeigt, dass das Lemma *rund* gleich in drei Wortformen vertreten ist: *runden Tisch*, *Runden Tisch* und *Runder Tisch*. Darüber hinaus treten diese drei Kollokationen überzufällig häufig zusammen mit weiteren Wörtern auf. Sowohl *runden Tisch* als auch



*Runder Tisch* sind mit *Gespräch* verbunden, erstere Kollokation auch mit der Pluralform *Gespräche(n)*.<sup>1</sup>

Neben dem Ausdruck *runden Tisch* gibt es zwei weitere wichtige Kollokationen: *Tisch ... Stühle* und *Tisch ... legen*. Diese beiden Kollokationen weisen wiederum je spezifische Verbindungen auf: Die Kette *Tische* → *Stühle* wird ergänzt durch die Bezeichnungen weiterer Möbelstücke: *Betten*, *Lampen* und *Schränke*. Die Kollokation *Tisch* → *legen* wird erweitert durch Bezeichnungen für Objekte, die auf den Tisch gelegt werden: *Schilling*, *Karten* und *Geld*.

Die Grafik basiert auf automatischen Berechnungen. Bei diesen Berechnungen werden Wortkombinationen extrahiert, die in einem Korpus häufiger auftreten, als es aufgrund der Frequenzen der einzelnen Wörter zu erwarten wäre. Dass zwei Wörter, die je sehr häufig in einem Korpus zu finden sind, auch oft zusammen im gleichen Satz auftreten, ist nicht überraschend. Überraschend ist hingegen die Kombination, wenn diese frequenter ist, als es bei einer zufälligen Verteilung der Wörter im Korpus der Fall wäre.<sup>2</sup>

Die Grafik zeigt, dass eine relativ simple statistische Auswertung von großen Textmengen zu Informationen über die Organisation von Sprachgebrauch führt. Einerseits geben die Berechnungen semantische Verbindungen wieder: Die Wörter *Tisch*, *Stühle*, *Betten*, *Schränke*, *Lampen* gehören alle zum selben semantischen Feld (MÖBEL). Andererseits werden sprachliche Repräsentationen von Praktiken wiedergegeben: Auf einen *Tisch* legt man *Geld* (manchmal auch *Schilling*) oder *Karten*. Und am *runden Tisch* führt man *Gespräche* oder man plädiert für die *Einrichtung* eines *Runden Tisches*.

Besonders letztere Beispiele zeigen, dass sich soziales Handeln sprachlich niederschlägt. Soziales Handeln führt zu einem typischen Sprachgebrauch, der statistisch auffällig ist. Es sollte also möglich sein,

<sup>1</sup> Die unterschiedlichen Grauwerte der Pfeile in Abbildung 1.1 auf der vorherigen Seite symbolisieren die Signifikanz der Verbindungen. Zu den Details der so berechneten Kollokationen vgl. Kapitel 6.

<sup>2</sup> Man mag sich wundern, weshalb *Schilling* als Kollokator zu *Tisch legen* erwähnt ist, zumal die Währung seit längerer Zeit nicht mehr existiert. An diesem Beispiel zeigt sich das Prinzip von Signifikanzberechnungen: Weil *Schilling* für sich gezählt keine sehr hohe Auftretensfrequenz aufweist, in überdurchschnittlich vielen Fällen aber zusammen mit *Tisch legen* erscheint, ist die Signifikanz der Verbindung auch sehr hoch. Offen bleibt jedoch, weshalb neben *Schilling* nicht auch *Mark*, *Franken* oder *Euro* als Kollokatoren erscheinen. Offensichtlich sind Verbindungen mit diesen Ausdrücken weniger signifikant, was natürlich auch an der Ausrichtung des Korpus liegen könnte.

von den Beobachtungen über typischen Sprachgebrauch in einem gewissen Maß auf die gesellschaftliche Organisation von Welt schließen zu können.

Ich habe oben auf die beiden orthographischen Varianten von *rund* hingewiesen. Es gibt also Belege für *Runder/Runden Tisch* und für *runden Tisch*. Eine detaillierte Analyse der Korpusbelege zeigt auf, dass *Runder Tisch* in der Großschreibung als Neulexem bezeichnet werden kann, das sich während der Wende Deutschlands 1989 gebildet hat (Herberg 1998, 50). Damit wird der Ausdruck idiomatisch: Bei einem *Runden Tisch* handelt es sich nicht bloß um einen Tisch, der rund ist, sondern um eine außerparlamentarische Institution von „Vertretern der verschiedenen politischen Richtungen nach polnischem und ungarischem Vorbild“ (Herberg 1998, 51).

Das Beispiel zeigt, wie idiomatische Ausdrücke besonders deutlich Abbild von gesellschaftlichen Institutionen sein können. Das Beispiel zeigt aber auch, dass bei diachroner Betrachtung von Sprachgebrauch dessen Veränderungen die Veränderungen sozialen Handelns abbilden. In dieser Perspektive tritt also zur Beobachtung, dass *Tisch* und *rund* überzufällig häufig zusammen verwendet werden, ein weiteres Kriterium: Diese Beobachtung ist typisch für eine bestimmte Zeit. Es wäre also zu erwarten, dass in Textkorpora, die Dokumente aus der Zeit der Wende enthalten, die Wortverbindung *Runder Tisch* nicht nur im Vergleich zu anderen möglichen Kombinationen von *Runder* und *Tisch* mit je anderen Wörtern besonders häufig vorkommen, sondern auch im Vergleich mit anderen Zeitabschnitten. Die Hypothese würde lauten: Es gibt eine musterhafte Verwendung von *Runder Tisch* während der Wende. Mit der Hypothese würde also historische Variation von Sprachgebrauch analysiert.

Dieses zweite Kriterium, die für einen bestimmten Sprachauschnitt typische Verwendung einer Wortverbindung, lässt sich selbstverständlich auch auf synchroner Achse anwenden: Eine Wortverbindung kann für eine bestimmte Textsorte, einen bestimmten Text, eine bestimmte kommunikative Form etc. typisch sein. Damit gilt die Analyse der synchronen Variation von Sprachgebrauch.

Nicht jede Wortverbindung, auch wenn sie idiomatisch ist, muss in einer zeitlichen Abhängigkeit stehen. In Abbildung 1.1 auf Seite 1 ist auch die Wortverbindung *Karten* → *legen* → *Tisch* aufgeführt,

die in den Belegen häufig in der Form *Karten (auf den) Tisch legen* verwendet wird. Es ist nicht zu vermuten, dass diese Wortverbindung typisch für eine bestimmte Zeitperiode ist – sie ist es vielleicht für bestimmte Textsorten.<sup>3</sup> Es könnte jedoch sein, dass diese Wortgruppe in Kombination mit bestimmten Kontexten zeitspezifisch verwendet wird. Der Blick müsste sich also auf die Verwendungskontexte richten, um *Karten (auf den) Tisch legen* als Indikator für soziales Handeln beschreiben zu können.

### 1.1 Eingrenzung

Ich habe bisher argumentiert, dass die Beobachtung von Veränderungen im Sprachgebrauch Rückschlüsse auf ‚soziales Handeln‘ zulässt. An dieser Stelle sind zwei wichtige Eingrenzungen nötig.

Alle Beobachtungen, die ich bisher beschrieben habe, betreffen nicht die Sprache, sondern den Sprachgebrauch. Es geht um spezifische Verwendungsweisen von Sprache, die spezifischen Präferenzen für eine mögliche Ausdrucksvariante unter anderen. Die Analysen, die ich anstrebe, haben nicht den Anspruch, ein etwaiges Sprachsystem zu beschreiben. Der Fokus liegt auf der Parole, nicht der Langue, um es mit Saussure’schen Termini auszudrücken. Damit soll nicht negiert werden, dass die Langue einen bestimmten Sprachgebrauch bedingen kann. Die Systematik der Langue zu erfassen, ist jedoch nicht das Ziel der vorliegenden Untersuchung.

Eine zweite Eingrenzung ist notwendig: Was ich oben grob als ‚soziales Handeln‘ beschrieben habe, kann ich in der vorliegenden Arbeit nicht als solches untersuchen. Stattdessen scheint mir der Diskursbegriff, und zwar im Foucault’schen Sinn, fruchtbar, um damit das sprachlich fassbare Produkt von sozialem Handeln zu fassen. Ich versuche also, über einen Umweg über Diskurs – oder besser: Diskurse – eine Verbindung zwischen Sprachgebrauch und sozialem Handeln herzustellen.

---

<sup>3</sup> Dass keine diachrone Variation erwartet werden kann, gilt allerdings nur für einen beschränkten, neueren Zeitraum. Überblickt man Zeiträume von mehreren Jahrzehnten oder gar Jahrhunderten würde sich auch bei dieser Wortverbindung Variation aufzeigen lassen.

Das Interessante am Diskursbegriff ist die Idee der ‚diskursiven Praxis‘. Sie ist das Ergebnis von Prozeduren der Produktion von Diskursen und widerspiegelt die getroffenen Selektionen von Sprechweisen: In einem bestimmten Diskurs wird etwas *so* ausgedrückt, obwohl es auch anders ausgedrückt werden könnte. Es muss deshalb das Ziel der vorliegenden Arbeit sein, diese typischen Sprechweisen über eine Analyse des musterhaften Sprachgebrauchs zu erfassen, um daraus Schlüsse über Diskurse zu ziehen.

Eine weitere Denktradition bietet sich an, um eine Verbindung zwischen Sprachgebrauch und sozialem Handeln herzustellen: In der Linguistik hat sich in den letzten Jahren eine vielfältige Forschungstätigkeit unter der Etikette ‚Kulturanalyse‘ entwickelt. Im Zentrum stehen oft diachrone Untersuchungen zu Veränderungen des Sprachgebrauchs, die Rückschlüsse auf kulturelle Praktiken ermöglichen. Ich möchte deshalb in der vorliegenden Arbeit alternativ zum Diskursbegriff auch kulturanalytische Konzepte wie Mentalitäten, diskurssemantische Grundfiguren, idiomatische Prägungen etc. miteinbeziehen und der Frage nachgehen, wie ‚Sprache lesen‘ und ‚Kultur lesen‘ zusammenhängen.

## 1.2 Hypothesenbildung

Musterhafter Sprachgebrauch zeigt sich in Wortverbindungen, die typisch für bestimmte Textmengen, also allgemeiner: Sprachausschnitte, sind. Diese Wortverbindungen üben eine Art Vorbildfunktion zur Produktion weiterer Wortverbindungen aus. Als Resultat ist dann ein musterhafter Sprachgebrauch in diesem Sprachausschnitt sichtbar.

Wie das Eingangsbeispiel gezeigt hat, kann musterhafter Sprachgebrauch statistisch operationalisiert werden. Es sollte also möglich sein, in einer großen Textmenge mit maschinellen Verfahren ‚Cluster‘ von typischem Sprachgebrauch zu extrahieren. Das Suchkriterium heißt also: Finde überzufällige Wortkombinationen, die musterhaft verwendet werden, also typisch für einen bestimmten Sprachausschnitt im Vergleich zu anderen Sprachausschnitten sind.

Eine so operationalisierte Suchstrategie erlaubt eine Methodik, die induktiv statt deduktiv vorgeht: Sie ermöglicht es, aus den Daten Mus-

ter zu extrahieren, ohne vorher im Detail definieren zu müssen, wie genau diese Muster aussehen. Es ist nicht nötig, eine Menge von Lexemen, Wortverbindungen oder syntaktischen Strukturen zu definieren und gezielt danach zu suchen. Mit diesem induktiven Vorgehen, das eine der grundlegenden Thesen, ergeben sich Vorteile für die Beantwortung diskurs- und kulturanalytischer Fragestellungen.

Zusammenfassend will ich die bisherigen Überlegungen zu folgenden Hypothesen zuspitzen, die in dieser Arbeit untersucht werden:

1. Sprachgebrauchsmuster sind Indikatoren für Diskurse.
2. Methodisch bietet sich damit die Chance, Sprachdaten auf ihre Musterhaftigkeit zu analysieren und daraus induktiv Diskursbeschreibungen abzuleiten.
3. Theoretisch ergibt sich daraus ein breiterer Begriff von Musterhaftigkeit, der sowohl Muster umfasst, die die Inhalte des Sprechens, als auch Muster, die das Wie des Sprechens, die ‚Sprechweise‘ ausmachen.

Das Ziel dieser Arbeit ist es, eine Methodik zu entwickeln, mit deren Hilfe es möglich ist, große Textkorpora nach typischen Mustern im Sprachgebrauch zu untersuchen. Diese Muster werden auf unterschiedlichen Ebenen zwischen Wort, Satz und Text auftauchen, und sie werden sowohl in einer synchronen als auch diachronen Perspektive betrachtet. Zudem werden die Sprachgebrauchsmuster sowohl thematisch/inhaltlich gefüllt (z. B. *Kampf gegen den Terrorismus*), als auch unabhängiger von bestimmten Themen sein, da sie thematisch/inhaltlich nicht gefüllt sind (z. B. *nicht nur ... sondern ... auch*). Die zu entwickelnde Methode nenne ich ‚korpuslinguistische Diskursanalyse‘.

In Teil I versuche ich zwischen sprachlicher Oberfläche und Tiefenstruktur Muster zu finden, die als Effekte von Sprachgebrauch angesehen werden können. Dafür muss auch die Beziehung zwischen Sprachgebrauchsmustern, Diskurs und Kultur geklärt werden.

Dieser Teil nimmt eine ‚Top-Down‘-Perspektive ein, da ich von theoretischen Überlegungen ausgehe und versuche, die unterschiedlichen Typen musterhaften Sprachgebrauchs mit linguistischen Konzepten zu beschreiben. Diese Konzepte befinden sich auf unterschiedli-

chen Niveaus zwischen Theorie und Phänomenbeschreibung und verfolgen unterschiedliche Ziele. Sie reichen von Stil über kommunikative Gattungen, Mentalitäten, idiomatischen Prägungen hin zum Topos- und Metaphernbegriff und Überlegungen einer Sprachgebrauchsgeschichte als Kulturanalyse. Bei jedem dieser Konzepte ist zu fragen, ob sie über die Analyse musterhaften Sprachgebrauchs erfasst und als Phänomen der Textoberfläche operationalisiert werden können. Denn das ist die Voraussetzung, um sie für korpuslinguistische Methoden nutzbar zu machen.

Die ‚Top-Down‘-Perspektive von Teil I muss mit einer ‚Bottom-Up‘-Perspektive in Teil II kontrastiert werden. Darin versuche ich aufzuzeigen, wie die Erkennung von Sprachgebrauchsmustern als Basis für Diskursanalysen dienen kann. Damit einher geht die Überzeugung, dass eine korpuslinguistische Sicht auf Sprachdaten Implikationen auf theoretischer Ebene zeitigt. Denn solche maschinellen Verfahren können induktiv angewendet werden, um musterhaften Sprachgebrauch in Sprachdaten sichtbar zu machen. Somit eröffnen diese Verfahren neue Möglichkeiten, um auf empirischer Basis Sprachgebrauch, und damit auch Diskurse, beschreiben zu können.

In Teil II gibt es zwei ausgesprochen technische Kapitel zu den statistischen Verfahren (Kapitel 7) und zu Ressourcen der Korpus- und Computerlinguistik (Kapitel 9), die in der vorliegenden Arbeit verwendet werden. Diese Kapitel richten sich primär an Leserinnen und Leser, die die statistischen und technischen Details nachvollziehen und ggf. für eigene Untersuchungen nutzen wollen.

In Teil III wird an einigen Beispielen demonstriert, wie eine korpuslinguistische Diskursanalyse funktionieren könnte. Untersuchungsgegenstand ist eine Zufallsauswahl von knapp 45 000 Zeitungsartikel der ‚Neuen Zürcher Zeitung‘ (NZZ-Korpus 1995–2005), die im Zeitraum von 1995 bis 2005 publiziert worden sind. Durch die Analyse, die zunächst typischen Sprachgebrauch in den Daten sichtbar macht, werden sich eine Reihe von Mustern ableiten lassen, die partiell genauer unter die Lupe genommen werden. Das Interesse liegt jedoch nicht primär in bestimmten Themen, sondern im typischen Sprachgebrauch einer Tageszeitung und dessen Wandel.

### 1.3 Terminologie und allgemeine Hinweise

In der vorliegenden Arbeit wird mit Termini argumentiert, deren Gebrauch im wissenschaftlichen Kontext unscharf ist. Ich möchte deshalb gleich zu Beginn die grundlegendsten Termini für die weitere Verwendung definieren, wobei ich mich grundsätzlich an die vorherrschenden Fachdefinitionen halte. Die Begründungen dieser Definitionen folgen teilweise erst in den entsprechenden Kapiteln. Sie seien aber bereits an dieser Stelle genannt, sodass diese Kurzdefinitionen schnell zur Hand sind.

*Wort:* Damit meine ich die konkret im Text vorkommenden Token. Dies vor allem im korpuslinguistischen Kontext, wo die Angabe „das Korpus enthält 152 000 Wörter“ bedeutet, dass im Korpus 152 000 Token vorkommen.

*Lexem:* Es bildet ein Paradigma verschiedener syntaktischer Wörter der gleichen Wortart mit demselben semantischen Merkmal (alle flektierten Formen *Haus*, *Häuser*, *Hauses*, *Hause* gehören zum Lexem *Haus*).

*Kookkurrenz:* Zwei Wörter, die nahe zusammen auftreten.

*Kollokation:* Zwei Wörter, die frequent und/oder überzufällig oft nahe zusammen in einem Korpus auftreten.

*Mehrworteinheit:* Zwei oder mehr Wörter, die frequent und/oder überzufällig oft nahe zusammen in einem Korpus auftreten.

*Muster:* Mehrworteinheiten, die einem interpretatorischen Selektionsprozess unterzogen und zu einer abstrakteren Einheit zusammengefasst wurden.<sup>4</sup>

*Sprachgebrauch und Sprechen:* Ich versuche den Terminus ‚Sprache‘ zu vermeiden und stattdessen vom ‚Sprachgebrauch‘ oder dem ‚Sprechen‘ zu sprechen. Damit sollen die pragmatischen, textoberflächlichen Aspekte betont und klar signalisiert werden, dass es nicht um die Untersuchung eines möglichen Sprachsystems geht,

<sup>4</sup> Vgl. für ‚Kookkurrenz‘, ‚Kollokation‘, ‚Mehrworteinheit‘ und ‚Muster‘ auch die Herleitung in Kapitel 6 und die Definition in Kapitel 6.3.

sondern um Sprachgebrauch, wie er sich auf der Textoberfläche in Korpora zeigt.

Darüber hinaus möchte ich auf Regeln zu Textauszeichnungen aufmerksam machen:

*Text- und Wortbeispiele* werden kursiv dargestellt.

Beispiel: Der Ausdruck *nicht nur* weist einen interessanten Sprachgebrauch auf.

Kursivierungen werden in seltenen Fällen auch als Auszeichnungsmittel für Betonungen im Fließtext verwendet.

*Sprachgebrauchsmuster und semantische Klassen* werden in Kapitälchen dargestellt.

Beispiel: Das Muster NICHT NUR ... SONDERN ... AUCH muss genauer analysiert werden.

*Metasprachliche Ausdrücke und Termini* werden in einfache Anführungsstriche gesetzt.

Beispiel: Der Begriff der ‚Kollokation‘ wird unterschiedlich definiert.

*Belege* werden vom Fließtext abgehoben, mit Zeilennummern und dem Verweis auf die Quelle versehen. Hervorhebungen in **fett** sind immer von mir.

Die Quellenangaben zu den Belegen sind je nach den verfügbaren Angaben der Originalquelle unterschiedlich detailliert aufgeführt. Immer vorhanden sind Titel und Datum, meistens auch Ressort und Autor/Autorin (manchmal als Kürzel) bzw. Agenturname.

Beispiel:

(1) Das ist ein Beleg. *Quelle*

*Mathematische Sonderzeichen* werden größtenteils in Kapitel 7 eingeführt. Normalerweise werden in den Analysen folgende Maße verwendet: % für relative, # für absolute Frequenzen;  $\odot$ : arithmetisches Mittel;  $\pm$  [Wert] oder s: Standardabweichung;<sup>5</sup>  $\chi^2$ : Wert des

<sup>5</sup> Vgl. Kapitel 7.1.



Chi-Quadrat-Tests; df: Freiheitsgrade der Kontingenztafel; p: Wahrscheinlichkeit, mit der die Nullhypothese  $H_0$  des Signifikanztests abgelehnt werden kann; V: Cramér's V.<sup>6</sup>

*Weiterführendes Material* zur vorliegenden Arbeit, wie ausführliche Typikprofile, Datensätze und verwendete Software, findet sich im Web unter der Adresse <http://www.bubenhofer.com/korpusanalyse/>.

---

<sup>6</sup> Vgl. zum Signifikanztest Kapitel 7.2.3.

# **I**

## **Muster als Effekte des Sprechens: Theoretische Perspektive**



Es ist das Ziel dieser Arbeit, einen Methodenapparat für eine korpuslinguistische Diskursanalyse zu entwickeln. In der Einleitung habe ich die Thesen dargelegt, auf denen ein solcher Methodenapparat beruht. Sie seien hier nochmals genannt:

1. Sprachgebrauchsmuster sind Indikatoren für Diskurse.
2. Methodisch bietet sich damit die Chance, Sprachdaten auf ihre Musterhaftigkeit zu analysieren und daraus induktiv Diskursbeschreibungen abzuleiten.
3. Theoretisch ergibt sich daraus ein breiterer Begriff von Musterhaftigkeit, der sowohl Muster umfasst, die die Inhalte des Sprechens, als auch Muster, die das Wie des Sprechens, die ‚Sprechweise‘ ausmachen.<sup>1</sup>

In diesem Teil wird der theoretische Hintergrund dieser Thesen dargelegt. So werden in Kapitel 2 die Grundbegriffe Sprachgebrauch, Muster und Diskurs geklärt. Diese Kernbegriffe werden um einen vierten Aspekt ergänzt: Kultur. Denn es wird sich zeigen, dass Diskursanalyse als Kulturanalyse gedacht werden kann, wenn Diskursanalyse mit einem stärkeren Fokus auf typische Sprechweisen ergänzt wird.

Anschliessend stellt sich in Kapitel 3 die Frage nach dem Zusammenhang von (musterhaftem) Sprachgebrauch und Diskursen. Weshalb ergibt sich überhaupt musterhafter Sprachgebrauch aus Diskursen?

Zuletzt werden in Kapitel 4 einige linguistische Konzepte dargestellt, die typische Sprechweisen zu erfassen versuchen und die nützlich sind, um die Musterhaftigkeit des Sprachgebrauchs als Phänomen auf der Textoberfläche beschreiben zu können.

Die These, dass eine induktiv ausgerichtete Korpuslinguistik das richtige Mittel ist, um die typischen Sprechweisen zu erfassen, wird hier nur indirekt begründet. Doch wird in Teil II genügend Raum dafür bleiben.

---

<sup>1</sup> Ich habe bereits in den terminologischen Klärungen (Kapitel 1.3) verdeutlicht, dass der Terminus ‚Sprechen‘ nicht etwa als Bezeichnung für ‚gesprochene Sprache‘, sondern als Betonung der pragmatischen Aspekte des Sprachgebrauchs verstanden werden soll.



## 2 Grundbegriffe

### 2.1 Sprachgebrauch korpuslinguistisch

Die Saussure'sche Dichotomie von *Langue* und *Parole*, von Sprachsystem und Sprachgebrauch, hat sich in der Linguistik als äußerst produktiv erwiesen. Dabei entfaltete sich diese Produktivität wechselweise aus der mehr oder minder strengen Beschränkung des Interesses auf jeweils eine Seite der Dichotomie. Die damit einhergehende Ignoranz der jeweils anderen Seite ist dann der Kern der Kritik von konkurrierenden wissenschaftlichen Konzepten.

So stellte z. B. die ‚pragmatische Wende‘, mit der die vorherrschende ‚Systemlinguistik‘ angegriffen wurde, den Sprachgebrauch ins Zentrum (vgl. weiter unten Seite 43). Und Wittgensteins Diktum, die Bedeutung eines Wortes liege in seinem Gebrauch (Wittgenstein 1995, 262; §43), stellte die Semantik auf eine neue pragmatische Basis.

Ähnliches gilt auch für alle Ansätze, die sich um die Schlagworte ‚linguistische Diskursanalyse‘, ‚Begriffsgeschichte‘ etc. versammeln: Untersuchungsobjekt ist die Sprache im konkreten Gebrauch, die sich in Zeitungsartikeln, Pamphleten, Flugblättern, Werbeanzeigen, Briefen oder SMS-Botschaften zeigt, um nur wenige Formen von Sprachgebrauch zu nennen. Werden z. B. Lexeme in ihrem Gebrauchskontext erfasst, ergibt sich eine semantische Beschreibung, die um pragmatische Aspekte ergänzt werden kann und damit komplexer wird, als wenn versucht wird, die Semantik von Einzellexemen kontextlos zu bestimmen. So verbergen sich beispielsweise hinter Ausdrücken des Migrationsdiskurses wie ‚Abschiebung‘, ‚Integration‘, ‚Missbrauch‘ oder ‚Asylant‘ mentale Konzepte wie *Topoi* oder Argumentationsfiguren in jeweils diskursspezifischer Ausprägung (vgl. Wengeler 2005, 42f.). Diese diskursspezifischen Funktionen können nur durch die Analyse des Sprachgebrauchs im je spezifischen Kontext beschrieben werden.

Der Paradigmenwechsel der pragmatischen Wende war zwar revolutionär, muss im Bereich der Diskurs- und Kulturanalyse jedoch heute

nicht mehr verteidigt werden. Trotzdem gibt es erneut Anlass, die alte Forderung nach dem Fokus auf den Sprachgebrauch zu diskutieren. Einen neuen Akzent setzt hier die (elektronische) Korpuslinguistik, von der noch immer unklar zu sein scheint, ob sie bloßes Hilfsmittel oder etwa doch mehr als das ist:

Corpus linguistics is not in itself a method: many different methods are used in processing and analysing corpus data. It is rather an insistence on working only with real language data taken from the discourse in a principled way and compiled into a corpus. However, one should be wary of using such data merely to find out more about what we know already, since what (we think) we know is often derived from pre-corpus study. Corpus data provide insights of a type which has not previously been available. (Teubert 2005, 4)

Die erste Forderung, prinzipiell nur mit ‚realen‘ Sprachdaten zu arbeiten, ist einsichtig und nicht neu. Doch wie kann behauptet werden, mit der Korpuslinguistik würden Einsichten ermöglicht, die vorher nicht möglich gewesen seien?

In der Linguistik wird schon seit geraumer Zeit mit Textkorpora, also Sammlungen von Text, gearbeitet. Was die moderne (elektronische) Korpuslinguistik jedoch davon unterscheidet, ist Folgendes:

1. In der Korpuslinguistik wird wann immer möglich mit großen Textmengen gearbeitet. Ein Korpus besteht nicht aus mehreren Texten, sondern eher aus mehreren Tausend Texten.
2. In der Korpuslinguistik liegt das Interesse nicht bei einzelnen Texten, sondern beim Sprachgebrauch in großen Textgruppen.
3. Damit wird die Korpuslinguistik als empirisch verstanden und es wird mit quantitativen Methoden gearbeitet.

Das ermöglicht eine andere Analyse von Sprache: Statt Sprachgebrauch in Einzeltexten zu beobachten, kann korpuslinguistisch Sprache als Datensammlung aufgefasst werden, in der es mit induktiven Verfahren möglich wird, Strukturen, die statistisch auffällig sind, sichtbar zu machen.<sup>1</sup>

---

<sup>1</sup> ‚Statistisch auffällig‘ bedeutet dabei auch, dass es sich um Auffälligkeiten handeln kann, die den Sprachbenutzern keineswegs bewusst auffallen, die aber dennoch signifikant sind.

Damit gibt es zwei unterschiedliche Zugänge zu Sprachgebrauch. Während nicht-korpuslinguistische<sup>2</sup> Arbeiten eher qualitativ und deduktiv vorgehen, sind korpuslinguistische Zugänge deutlich quantitativ und eher induktiv angelegt. Was ist mit ‚deduktiv‘ und ‚induktiv‘ gemeint?

Mit dem induktiven Vorgehen versucht man aus der Beobachtung der Daten Regeln abzuleiten. Das können Regeln der Verteilung von sprachlichen Einheiten im Korpus, z. B. in Abhängigkeit einer anderen Variable wie Textart, Zeit oder Position im Text sein. Wichtig ist hierbei zu vermeiden, voreilig mit linguistischen Konzepten die Daten vorzustrukturieren. Stattdessen ist es das Ziel, die linguistische Kategorienbildung strikt aus den empirischen Beobachtungen herzuleiten.

Bei einem deduktiven Vorgehen dagegen ist schon vor der Korpusrecherche möglichst exakt definiert, wonach gesucht werden soll und welche Kategorien als relevant gelten. Hier ist es wenig wahrscheinlich, auf bislang unbekannte oder unerwartete Phänomene zu stoßen, was auch nicht das vordringliche Ziel ist. Im Zentrum steht die Bemühung, eine klar definierte Hypothese anhand der Korpusrecherche zu überprüfen.

In der Forschungspraxis ist ein Entweder-oder von induktivem und deduktivem Vorgehen nicht sinnvoll. Induktiv abgeleitete Regeln müssen sich bewähren, indem sie deduktiv an anderen Daten getestet werden.

Aus korpuslinguistischer Perspektive werden diese beiden Zugänge als ‚corpus-driven‘ und ‚corpus-based‘ bezeichnet. Diese beiden Perspektiven werden in Kapitel 5.1 weiter ausgeführt.

Die pragmatische Wende hat in der Linguistik ein neues Interesse für die Analyse von Sprachgebrauch ausgelöst. Mit Hilfe der Korpuslinguistik, die mit quantitativen Methoden eine große Datenbasis verarbeitet, ergeben sich ganz neue methodische Hilfsmittel. Eine verstärkt empirisch-quantitativ arbeitende, korpuslinguistische Sprach-

---

<sup>2</sup> Mit Recht können Forschende auch ihre sorgfältige Auswahl von 100 Texten als Korpus bezeichnen oder darauf verweisen, dass bereits das Grimmsche Wörterbuch (Grimm 1885) auf Basis eines Textkorpus erstellt worden sei. Trotzdem möchte ich solche Arbeiten mit relativ kleinen Korpora nicht als ‚korpuslinguistisch‘ bezeichnen, sondern verstehe unter dem Begriff die quantitativ-empirische Arbeit mit elektronischen Textkorpora.



gebrauchsanalyse vereinfacht es, aus großen Datenmengen induktiv Strukturen herauszuarbeiten und daraus ggf. neue linguistische Kategorien abzuleiten. Denn mit statistischen Verfahren der Korpuslinguistik können typische Sprachgebräuche sichtbar gemacht werden, die durch eine Lektüre der Texte nicht erkannt würden, sei es, weil die zu überblickende Datenmenge zu groß ist, sei es, weil diese Sprachgebräuche zwar statistisch auffällig sind, sich aber der bewussten Aufmerksamkeit der Forscherin oder des Lesers entziehen.

## 2.2 Muster

Muster im Sprachgebrauch zu entdecken ist die zentrale Operation einer korpuslinguistischen Diskursanalyse; so die bereits in Kapitel 1 erarbeitete These. Mit den Begriffen Muster und Musterhaftigkeit klingen Konzepte an, die in der Linguistik bereits mit anderen Begriffen, wie etwa ‚Type/Token‘, ‚Typik‘, ‚Schema‘, ‚Prototyp‘ oder ‚Regel‘ gefasst werden. Sie decken sich jedoch nur teilweise mit der Idee des Musters, wie die folgenden Ausführungen zeigen werden.

### 2.2.1 Bedeutungsvielfalt

Das Wort ‚Muster‘ geht auf das Italienische ‚mostra‘ zurück, das mit ‚Probestück‘ bzw. ‚Ausstellung‘, ‚Auslage‘ übersetzt werden kann und seinerseits auf lat. ‚monstrare‘ (zeigen) zurückgeht (Kluge 1995, ‚Muster‘). Das Probestück bewegt sich dabei zwischen zwei Polen: Einerseits ist es das erste erstellte Objekt, das vorher nur als Plan existierte. Andererseits ist es Demonstrationsobjekt, das zeigt, wie die kommenden noch herzustellenden Objekte aussehen werden. In diesem Sinne ist es die Vorlage für die zu erstellenden Objekte. (Genau so verhält es sich bei der Produktion von Ravioli: Der erste Prototyp der Ravioli, den ich aus dem Teig geschnitten, mit Füllung versehen und zusammengefaltet habe, erfüllt beide Funktionen: Er dient als Versuch, um zu sehen, ob die Ausführungen im Rezept [= der Plan] funktionieren, gleichzeitig als Vorlage für die weiter zu erstellenden Ravioli. So kann ich zu meiner Mitköchin sagen: Schau, so sollen sie am Schluss aussehen!)

Interessant ist bei dieser Verwendungsweise von ‚Muster‘ Folgendes: Das Probestück ist (mehr oder weniger) von gleicher Klasse wie die in der Folge sich daran orientierenden ‚endgültigen‘ Objekte. Es handelt sich somit beim Probestück nicht um den Plan, sondern um ein Objekt, dem der Status der Vorlage zuerkannt wird. Diese Abgrenzung des Musters gegenüber dem Plan ist wichtig, wie ich später zeigen möchte.

Doch neben dieser Bedeutung von ‚Probestück‘ gibt es weitere Bedeutungen. Der Duden (1999) nennt vier Bedeutungsaspekte, die vom eben referierten mehr oder weniger differieren:

1. *Vorlage, Zeichnung, nach der etwas hergestellt, gemacht wird*: etwas dient als Muster; ein Kleid nach einem Muster schneiden; Ü<sup>3</sup> ein Justizwesen nach angelsächsischem Muster; Der Stern der nach eingefahrenen Mustern gepflegten Politik ist offenbar im Sinken (Brückenbauer 11.9.85, 1).
2. *etwas in seiner Art Vollkommenes, nachahmenswertes, beispielhaftes Vorbild in bezug auf etwas Bestimmtes*: sie war ein Muster an Geduld, Fleiß; er ist das Muster eines guten Vaters; mein Café ist ein Muster von Gepflegtheit und Sauberkeit (Hamburger Morgenpost 5.9.84, 9); jemanden zum Muster nehmen.
3. *aus der Kombination von einzelnen Motiven bestehende [regelmäßige], sich wiederholende, flächige Verzierung, Zeichnung auf Papier, Stoff o.ä.*: ein großes, buntes, auffallendes Muster; das Muster einer Tapete, eines Stoffes; ein Muster entwerfen, zeichnen, stricken; den Pullover in einem anderen Muster stricken; Ü das läuft hier doch immer nach dem gleichen Muster (*Schema*) ab.
4. *kleines Stück, kleine Menge einer Ware, an der man die Beschaffenheit des Ganzen erkennen kann*: Muster von Stoffen, Tapeten, Wolle; Muster anfordern; sie ließen sich von dem Vertreter die neuesten Muster zeigen; Muster ohne Wert (Postwesen veraltend; *Warensendung*).

(Duden 1999, ‚Muster‘)<sup>4</sup>

Die Bedeutung des ‚Probestücks‘, wie ich sie oben entwickelt habe, findet sich in diesen Definitionen nur teilweise wieder: Das Probestück dient als Vorlage, wie in 1 beschrieben, jedoch ohne erst Zeichnung zu sein; es ist bereits ein gefertigtes Stück. Bedeutungsaspekt 2 scheint ein Spezialfall von 1 zu sein: Diese Muster sind ebenfalls Vorlagen, allerdings zwangsweise positiver Art. Interessanterweise sind hier die aufgeführten Beispiele allesamt kompatibel mit der Bedeutung von ‚Probestück‘, da die ‚Objekte‘ (geduldige, fleißige Person, guter Vater,

<sup>3</sup> Ü steht für ‚Übertragung‘, NB.

<sup>4</sup> Die Darstellung wurde der besseren Lesbarkeit wegen leicht abgeändert.

Café) einerseits als Vorbild dienen, andererseits auch gleicher ‚Klasse‘ sind, wie die nach ihnen entstandenen ‚Abbilder‘.

Mit Bedeutungsaspekt 3 hat die Bedeutung von ‚Probestück‘ hingegen nichts zu tun; mit Aspekt 4 zumindest partiell im Sinne, dass das Probestück eine „kleine Menge einer Ware“ ist. Allerdings ist das Probestück im Gegensatz zu diesem Bedeutungsaspekt vollständig und nicht nur ein Teil davon. Es ist offensichtlich: Die etymologische Bedeutung von ‚Muster‘ differiert von der nach Duden heute gebräuchlichen.

Die Definitionen von ‚Muster‘ im Wörterbuch der Gebrüder Grimm (Grimm 1885) sind mit gut drei Spalten ausführlicher und doch im Großen und Ganzen ähnlich der Duden-Definition. Gerafft, und um die meisten Beispiele gekürzt, zeigt sich die Struktur folgendermaßen:

MUSTER, *n. was man zeigt, probestück, vorbild u.a.*

1. *das wort, seit dem 15. jahrh. bei uns eingebürgert, ist lehnwort aus dem romanischen, ital. mostra, franz. monstre, später montre, in welchen ländern es unter manchen andern bedeutungen in gewerblichen kreisen auch die eines zur schau und zur probe vorgezeigten stückes, einer kunstgewerblichen arbeit, nach der man andere liefern konnte, hatte. [...]*
2. *muster, als gebliebenes wort der gewerblichen und der kaufmännischen sprache, in mehrfacher ausbildung eines begriffs.*
  - a) *probestück, vorlage für etwas danach zu fertigendes [...]*
  - b) *später besonders im gebiete des kleingewerbes: muster, nennen die handwerksleute ein vor sich habendes modell, abrisz, probe, darnach sie ihr ganzes werk zu machen haben. [...]*
  - c) *gärtner nennen muster das anlegen einer zierlichen figur im parterre des blumengartens [...]*
  - d) *muster auch ein von einer grösseren menge entnommenes stück oder kleinen theil [...]*
  - e) *muster überträgt sich auf die zeichnung, die ein solches stück zeug oder eine stickerei und ähnl. nach vorausgemachtem entwurf (oben b) empfangen hat: der kaufmann ... weisz aus der erfahrung, dasz, wenn ihr euch lange und viel besonnen habt, ihr endlich doch auf das schlechteste fallt; auf eine farbe, auf ein muster, das längst nicht mehr mode gewesen. LESSING I, 358; [...] von tapeten: tapeten mit einfachen mustern; von schablonierten tellern, tassen u.a.: die Meiszner teller mit dem hässlichen zwiebel-muster.*
3. *muster, in die allgemeine sprache übergegangen, und zwar als edler ausdruck.*

- a) *von dingen, vorbild, nach dem man sich richtet [...]*
- b) *auch probe (vergl. 2, d), zur erkennung des wesens eines dinges [...]*
- c) *dann von personen, deren art als vorbild dient. diesselbe ist durch begleitenden genitiv bezeichnet [...]*

4. *muster, ironisch, von einem schlechten dinge [...]*

(Grimm 1885, „MUSTER“)<sup>5</sup>

Wir sehen an diesen Definitionen, dass hier die etymologische Bedeutung (Muster als Probestück) noch sehr prominent dargestellt wird, dafür übertragene Bedeutungen nur ansatzweise vorhanden sind (Bedeutungsaspekt 3c). Die Idee des schematischen Prozesses (gemäss Duden-Beispiel „das läuft hier doch immer nach dem gleichen Muster“, Bedeutungsaspekt 3) ist bei Grimm noch nicht enthalten.

In einer der englischen Entsprechungen zu ‚Muster‘, nämlich ‚pattern‘, fällt ein weiterer Bedeutungsaspekt auf, den die Duden-Definition des deutschen Wortes ‚Muster‘ nur teilweise abgedeckt:

pattern [...] a way in which sth happens, moves, develops or is arranged: *patterns of behaviour/behaviour patterns* ◦ *the patterns of worldwide economic decline* ◦ *These sentences all have the same grammatical pattern.* ◦ *The murders all seem to follow a set/similar pattern* (ie occur in a similar way). (Oxford Advanced Learner's Dictionary 1995, ‚pattern‘)

In Teilen entspricht diese Definition dem 3. (übertragenen) Bedeutungsaspekt in der Definition gemäß Duden, wobei dieser dort als Synonym das ‚Schema‘ nennt, das wiederum nicht die gesamte Bedeutung der englischen Definition oben abdeckt. Denn diese Definition betont bei ‚pattern‘ nicht bloß einen schematischen Vorgang („sth happens, moves, develops“), sondern auch ein ‚pattern‘, das einer regelmäßigen Struktur („sth is arranged“) entspricht – und das nicht im engen Sinn („aus der Kombination von einzelnen Motiven bestehende [regelmäßige], sich wiederholende, flächige Verzierung, Zeichnung auf Papier, Stoff o.ä.“), sondern auch im weiteren Sinn, wie beispielsweise in: *These sentences all have the same grammatical pattern.*

### 2.2.2 Klärung: Beispiele und Definition

Es gilt nun, einen Musterbegriff zu finden, womit ich in der vorliegenden Arbeit die mich interessierenden Aspekte abdecken kann. Und

<sup>5</sup> Die Darstellung wurde der besseren Lesbarkeit wegen leicht abgeändert.

dieser Begriff liegt nahe an der zu Beginn dieses Kapitels genannten etymologischen Bedeutung, wie ich zeigen werde. Doch zunächst betrachte ich Phänomene, die unter dem Musterbegriff gefasst werden sollen:

1. Die Wortfolge *Krieg gegen den Terror(ismus)* tritt in einem untersuchten Korpus A mit einer Frequenz von x auf.
2. Die Wahrscheinlichkeit, dass das Wort *Mal* in einem Korpus B zusammen mit den Wörtern *zum* und *ersten* in einer Entfernung von höchstens fünf Wörtern auftaucht, ist signifikant höher, als es der Zufall voraussagen würde.
3. Wenn in einem Korpus C die Phrase *es liegt in der Natur der Sache* auftaucht, ist das in den meisten Fällen ein Teil einer typischen Argumentationsfigur, die eine Prämisse als unbestreitbar darstellen möchte.

Welche Aspekte an diesen Beispielen sind es nun, die als ‚Muster‘ gefasst werden sollen?

Die Wortfolge *Krieg gegen den Terror(ismus)* aus Beispiel 1 ist eine Verbindung von Wortformen, die in genau dieser Kombination oft vorkommt. Daneben werden die einzelnen Wortformen auch in Kombination mit anderen Wortformen verwendet. Doch da diese Kombination besonders häufig ist, möchte ich hier von einem ‚musterhaften Sprachgebrauch‘ sprechen. Das Muster *Krieg gegen den Terror(ismus)* wurde zur Vorlage, die oft kopiert, bzw. wiederholt wird. Es gab einen Plan, der die Produktion der Phrase ermöglichte: Regeln der Syntax, der morphosyntaktischen Flexion, der Semantik und der Pragmatik.

Bei Phänomen 2 handelt es sich um einen ähnlichen Fall. Der Unterschied liegt in den genaueren Angaben zur Typik des Musters. ‚Muster‘ möchte ich auch hier die häufige (und signifikante) Verbindung des syntaktischen Wortes *Mal* mit *zum* und *ersten* nennen.

Im 3. Phänomen kommt eine ‚typische Argumentationsfigur‘ ins Spiel. Die Phrase *es liegt in der Natur der Sache* ist nur ein verbalisierter Teil dieser Argumentationsfigur. Die ganze, aber abstrakte,

nicht verbalisierte Argumentationsfigur<sup>6</sup> führte zur Produktion dieser Phrase. Die Phrase *es liegt...* selbst, aber auch die daraus abgeleitete abstrakte Argumentationsfigur, kann nun als Vorlage für die weitere Produktion dieser Argumentationsfigur in konkreten Äußerungen dienen.

Ich fasse die Eigenschaften zusammen, die an diesen Beispielen meinen Musterbegriff geformt haben:

Ein (sprachliches) Muster

1. ist eine Wortform, eine Verbindung von Wortformen oder eine Kombination von Wortformen und nichtsprachlichen Elementen, also ein Zeichenkomplex,
2. der als Vorlage für die Produktion weiterer Zeichenkomplexe dient,
3. dabei aber von gleicher Materialität ist, wie die daraus entstehenden Zeichenkomplexe.

Diese Definition deckt sich zu einem großen Teil mit der Idee des ‚Probestücks‘, die sich aus der etymologischen Bedeutung von ‚Muster‘ (nämlich it. ‚mostra‘) ableitet. Gleichzeitig ist der Aspekt der ‚Vorlage‘, wie in Bedeutung 1 der Duden-Definition („Vorlage, Zeichnung, nach der etwas hergestellt, gemacht wird“), in der Definition oben enthalten, allerdings nicht im Sinne eines Types, sondern nur als Token, das selbst schon den daraus abgeleiteten Elementen gleichwertig ist. (Das Vorlage-‚Raviolo‘ ist bereits ein Raviolo, das nachher ebenfalls verspeist wird, nicht bloß die Idee davon. Und jedes Exemplar der danach produzierten Ravioli hat, sofern es hinreichend ähnlich wie die Vorlage ist, das Potential Vorlage-Raviolo zu werden.) Das stellt also eine wichtige Komponente von Mustern dar: Etwas wird zu einem Muster im Sinne einer Vorlage *gemacht* und ist es nicht per se. Jeder Zeichenkomplex kann in einer bestimmten Situation die Funktion eines Musters übernehmen. Außerhalb dieser Situation kann aber nicht mehr darüber entschieden werden, ob das Objekt diese Funktion je inne hatte – genau wie das (zumindest bei den meisten) sprachlichen

---

6 Diese Argumentationsfigur könnte z. B. mit der Schlussregel SCHLUSS VOM FAKTISCHEN AUF DAS NORMATIVE gefasst werden. Vgl. dazu auch Kapitel 4.5.

Mustern ebenfalls der Fall ist: Wenn vom *Kampf gegen den Terrorismus* die Rede ist, kann ein ursprüngliches Muster, das Vorbild für diesen Ausdruck war, nicht mehr ausgemacht werden. Trotzdem werden in der Sprachproduktion immer wieder Instanzen des Ausdrucks *Kampf gegen den Terrorismus* zu Mustern gemacht und führen so zu typischem bzw. musterhaftem Sprachgebrauch.

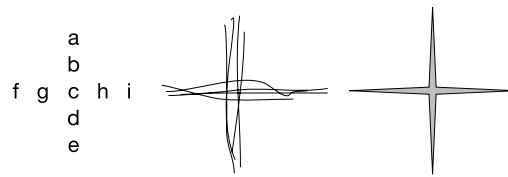
Anders gewendet: Ein Muster kann nur auf einer analytischen Ebene im Nachhinein festgestellt werden. Auf der Ebene des Sprachgebrauchs ist diese Musterfunktion für die Sprecherinnen und Sprecher kaum sichtbar. Mit ‚musterhafter Sprachgebrauch‘ wird deshalb betont, dass anscheinend im untersuchten Sprachausschnitt immer wieder Instanzen einer bestimmten Phrase als Muster (als Vorbilder) für die Produktion weiterer Instanzen dienten. Im Nachhinein ist aber nicht mehr erkennbar, welche Instanzen je diese Musterfunktion übernahmen. Aber der Effekt dieser unzähligen Instanzen, die einerseits einem Muster folgten und andererseits Musterfunktion übernahmen, ist auf der Ebene der Analyse als Phänomen eines typischen, oder eben: musterhaften Sprachgebrauchs sichtbar.

Hier kann eingewandt werden, dass für einen bestimmten Ausdruck sehr wohl oft ein Erstbeleg nachweisbar ist, dem wiederum Vorbildfunktion für weitere Verwendungen zugeschrieben werden kann. Das ist jedoch eine Ex-post-Interpretation, bei der dieser Verwendung durch den Schritt der Analyse eine Musterfunktion zugeschrieben wird. Im direkt an die Erstverwendung anschließenden Sprachgebrauch wäre diese Musterfunktion wahrscheinlich nicht erkannt worden.

Ich möchte den Musterbegriff nicht in Bezug auf die Detailliertheit der Definition einschränken. Die Instanzen, die Musterfunktion aufweisen, können Phrasen wie das oben Erwähnte *Kampf gegen den Terrorismus* sein, die bezüglich verwendeter Wortformen und lexikalischen Füllungen genau definiert sind, oder aber auch Phrasen wie KAMPF GEGEN X, die Slots für variable Füllungen offen halten.

### 2.2.3 Muster als Muster erkennen

In der erarbeiteten Definition oben habe ich betont, dass Zeichenkomplexe eine Funktion als Muster erfüllen können, aber nicht im ontologischen Sinne Muster sind. Dies bedeutet, dass ein bestimmter



**Abbildung 2.1:** Muster bestehen aus Ensemblestücken (Abbildung NB nach Savigny 1994, 18).

Zeichenkomplex als Muster verwendet werden kann, aber nicht muss. Da stellt sich die Frage, ob Bedingungen definiert werden können, die es begünstigen, dass ein Zeichenkomplex als Muster wahrgenommen wird.

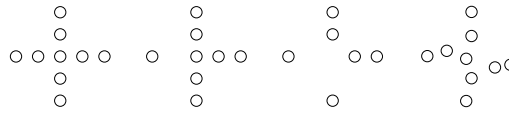
In Wittgensteins „Philosophischen Untersuchungen“ (Wittgenstein 1995) spielt der Musterbegriff eine tragende Rolle. Allerdings ist dieses Konzept, das man als „Musterrezept“ (Savigny 1994, 17) bezeichnen könnte, nirgends zusammenhängend dargestellt. Aber der Verfasser eines Kommentars zu den Philosophischen Untersuchungen, Savigny (1994, 17–27), versucht Wittgensteins Musterrezept zusammenzufassen.

Wittgensteins Konzept ist für meine Zwecke interessant, da es den Prozess der Erkennung von Zeichenkomplexen – oder allgemeiner: von Objektkonstellationen – genauer beschreibt. Allerdings wird bei Wittgenstein „Muster“ [...] in dem alltäglichen Sinne gebraucht, den es in ‚gemustert‘ hat (nicht im Sinne von ‚Vorbild‘ oder ‚Vergleichsmaßstab‘)“ (Savigny 1994, 17). Trotzdem scheinen mir die Überlegungen auch für meinen Musterbegriff hilfreich.

Savigny (1994, 17–27) führt – und nicht Wittgenstein, zumindest nicht explizit – neben ‚Muster‘ die Bezeichnungen ‚Ensemble‘ und ‚Ensemblestücke‘ ein, sowie ‚Mustersachverhalte‘ und ‚Ensemblesachverhalte‘.<sup>7</sup> Die Abbildung 2.1 zeigt, was damit gemeint ist. Das, was als Kreuz erscheint, ist ein Ensemble. Es besteht aus Ensemblestücken, also den einzelnen Buchstaben, Strichen oder Formen. Im Ensemble ist ein Muster zu erkennen, nämlich ein mehr oder minder ausgeprägtes Kreuz.

<sup>7</sup> Ich folge hier in geraffter Form Savigny (1994, 17–27).





**Abbildung 2.2:** Unterschiedlich starke Ausprägungen von Mustern (Abbildung NB nach Savigny 1994, 19).

Das Ensemble ist nicht dasselbe wie das Muster. Den Sachverhalt des Ensembles in der ersten Grafik in Abbildung 2.1 auf der vorherigen Seite kann man z. B. so erklären: „a steht über b, b steht über c, h steht links von i etc.“ Den Sachverhalt des Musters hingegen würde man beispielsweise so erklären: „a – i bilden ein Kreuz, wobei a, f, e und i die Enden sind, c der Mittelpunkt des Kreuzes.“ Der Unterschied der beiden Beschreibungen liegt im Wissen um die Existenz eines Musters: Den Sachverhalt eines Ensembles kann man auch beschreiben, wenn man das Muster ‚Kreuz‘ nicht kennt. Es handelt sich einfach um eine Beschreibung der Positionen der Buchstaben. Im zweiten Fall ist es aber zwingend, dass vor der Beschreibung das Ensemble als Muster eines Kreuzes erkannt wurde. Erst dann können die Buchstaben a, f, e und i als Enden und c als Mittelpunkt eines Kreuzes definiert werden.

Wenn man die Position und Funktion der Ensemblestücke in Bezug auf das Ensemble erklären möchte, bewegt man sich auf einer anderen Beschreibungsebene, wie wenn dies in Bezug auf das Muster gemacht wird. So würde man im Zusammenhang mit dem Muster ‚Schachbrett‘ einmal von ‚(Schachbrett-)Feldern‘ sprechen, aus denen das Muster besteht, ein anderes Mal von ‚Quadraten‘, die das Ensemble ausmachen.

Natürlich können Ensembleelemente nicht nur räumlich angeordnet sein, um ein Muster zu bilden, sondern auch zeitlich. So kann ein Tanz als Muster angesehen werden, wobei die Ensembleelemente (die Tänzerinnen und Tänzer) nicht nur selber bestimmte Positionen einnehmen und Handlungen (Schritte) durchführen müssen, sondern auch untereinander sich auf eine bestimmte Weise verhalten müssen.

Ein Muster kann unterschiedlich stark ausgeprägt sein. Das Muster ist kaum oder nicht ausgeprägt, wenn zu viele Ensembleelemente fehlen oder am falschen Platz sind, wie Abbildung 2.2 zeigt. Dabei ist unbestimmbar, an welchen Ensemblestücken es liegt, dass das Muster

nicht hinreichend ersichtlich ist. Denn 1) das gleiche Muster kann durch komplett verschiedene Ensemblestücke gebildet werden und 2) die Ausprägung eines Musters ist nicht nur von den Ensemblestücken abhängig, sondern auch vom Kontext. In Abbildung 2.2 auf der vorherigen Seite wird die dritte Figur im Kontext der ersten zwei Figuren ausgeprägter als Kreuz wahrgenommen, als wenn sie alleine stünde.

Die Termini ‚Ensemble‘, ‚Ensemblestücke‘ und ‚-sachverhalte‘ helfen, das Verständnis von Muster noch zu verfeinern. Denn bei sprachlichen Zeichenkomplexen können die einzelnen sprachlichen Elemente beispielsweise als Ensemblestücke eines Ensembles (KRIEG GEGEN TERROR[ISMUS]) verstanden werden, das als Muster erkannt werden kann, und zwar dann, wenn diesem Zeichenkomplex die Funktion eines Musters als Vorbild zuerkannt wird. Wichtigste Bedingung dafür, ob es als Muster erkannt wird (und nicht einfach als Ensemble), ist sein umgebender Kontext. So wird wahrscheinlich die Wortfolge *Krieg gegen Terror* in der Zeitungsberichterstattung oder in einer politischen Rede im Nachgang zum 11. September 2001 eher in der Funktion eines Musters verwendet als davor. Dies können Frequenzanalysen aufzeigen, wenn sie belegen, dass nach dem 11. September 2001 die Formulierung *Krieg gegen Terror* in den erwähnten Kontexten zunimmt, und zwar auf Kosten alternativer Formulierungsmöglichkeiten wie *den Terrorismus bekämpfen*, *Kampf gegen Terror*, *terroristische Netzwerke verfolgen* und dergleichen mehr.

Der Kontext bestimmt aber auch mit, welche Variationen des Ensembles möglich sind, so dass es noch immer dem selben Muster zugeordnet wird. So müssen Wortfolgen wie *Krieg gegen Terror*, *totaler Krieg gegen den Terror*, *Massnahmen gegen den Terror* oder *internationaler Kampf gegen den Terror* nicht zwangsweise als Ausprägungen desselben Musters verstanden werden. Die folgenden Belege illustrieren die unterschiedlichen Verwendungsweisen dieser Wortfolge:

- (2) Ministerpräsident Peres kündete einen **totalen Krieg gegen den Terrorismus** an. *Neue Zürcher Zeitung* vom 4. März 1996, Ressort ‚Ausland‘, Associated Press: „Neuer Anschlag auf einen Autobus in Jerusalem Tat eines Selbstmordattentäters – mindestens 19 Tote“.

- (3) Die G-7-Politiker verabschiedeten auf das Verlangen von Clinton noch am ersten Abend eine **Erklärung gegen den Terrorismus**. *Neue Zürcher Zeitung* vom 28. Juni 1996, Ressort ‚Ausland‘, Meister U.: „Politisch befrachteter G-7-Gipfel in Lyon. Für und wider die Globalisierung der Wirtschaft“.
- (4) Die von Spanien gelobte **Zusammenarbeit gegen den Terrorismus** ist einigen Schwankungen unterworfen. *Neue Zürcher Zeitung* vom 5. September 1998, Ressort ‚Ausland‘, Meister U.: „Wahlplattform der radikalen Basken. Mehrdeutiger ‚Coup‘ von Herri Batasuna“.
- (5) Der jugoslawische Präsident Kostunica hat **Massnahmen gegen den Terrorismus** angekündigt. *Neue Zürcher Zeitung* vom 20. Februar 2001, Ressort ‚Ausland‘, Wysling A.: „Provokationen und Kämpfe in Südserbien. Schuldzuweisungen nach den jüngsten Anschlägen“.
- (6) Einer aktiven Rolle in **Bushs Krieg gegen den Terror** stehen zudem hohe innenpolitische Hürden entgegen. *Neue Zürcher Zeitung* vom 26. September 2001, Ressort ‚Ausland‘, Schoettli U.: „Mühe Tokios mit Bushs Anti-Terror-Kampagne. Hohe innenpolitische Hürden gegen eine aktive Rolle“.
- (7) Putin unterstrich die Bereitschaft des Kremls zur Kooperation im **internationalen Kampf gegen den Terrorismus** und signalisierte sein Einverständnis mit einer allfälligen amerikanischen Militäroperation. *Neue Zürcher Zeitung* vom 26. September 2001, Ressort ‚Ausland‘, Gujer E.: „Deutsch-russischer Antiterror-Gipfel. Rede Präsident Putins vor dem Bundestag in Berlin“.

Die Wortfolge *gegen den Terror* erscheint in völlig unterschiedlichen politischen Kontexten und weist verschiedene Denotationen und Konnotationen auf.

Gleichzeitig finden sich Zeitungsartikel im NZZ-Korpus, in denen Belege für mehrere Varianten von *gegen den Terror* zu finden sind und diese Wortfolgen zumindest dieselbe Denotation aufweisen. Die Ensembles *totaler Krieg gegen den Terrorismus* und *Kampf gegen den*

*Terrorismus* haben so im folgenden Textausschnitt – in diesem Kontext – das Potenzial, als Varianten desselben Musters zu funktionieren:

- (8) Ministerpräsident Peres kündete einen **totalen Krieg gegen den Terrorismus** an. Präsident Weizman rief zu einer sofortigen Unterbrechung der Friedensverhandlungen auf. Die Friedensgespräche mit Syrien wurden auf unbestimmte Zeit ausgesetzt. [...] In einer Pressekonferenz nach der Kabinettsitzung erklärte Ministerpräsident Peres mit einer von ihm nicht gewohnten Härte und Entschlossenheit, dass der **Kampf gegen den Terrorismus** fortan oberste Priorität erhalten werde. 800 Sicherheitsleute sollen die Autobusse sichern und die Passagiere vor dem Einsteigen kontrollieren. *Neue Zürcher Zeitung vom 4. März 1996, Ressort ‚Ausland‘, Associated Press: „Neuer Anschlag auf einen Autobus in Jerusalem. Tat eines Selbstmordattentäters – mindestens 19 Tote“.*

Im Gegensatz zu den Belegen 2–7, die immer wieder unterschiedliche Kontexte der Ausprägungen von *gegen den Terror* aufweisen und diese deshalb eher nicht als Instanzen desselben Musters aufgefasst werden können, ist das in Beleg 8 anders. Die beiden Ausprägungen *totaler Krieg gegen den Terrorismus* und *Kampf gegen den Terrorismus* mögen zwar unterschiedliche Konnotationen aufweisen, werden jedoch eher als Instanzen desselben Musters aufgefasst.

#### 2.2.4 Abgrenzungen und Fazit

Meine Definition von ‚Muster‘ muss nun gegenüber anderen, ähnlichen Definitionen abgegrenzt werden.

Als ‚Schema‘ verstanden werden Formen der „Repräsentation von generalisiertem, soziokulturell bestimme[m] Wissen, das als Orientierung bei der Interpretation und zur Organisation von Erfahrungen dient“ (Bußmann 2002, ‚Schema‘). Es handelt sich also um eine komplexe Art von ‚Muster‘, die für meine Zwecke zu einschränkend wirkte.

Der Type in der ‚Type-Token‘-Relation wird als abstrakte Größe verstanden, die selber nicht gleichzeitig Token sein kann. Doch genau dies möchte ich mit dem Musterbegriff betonen: Dass ein Token

oder eine Token-Kombination die Funktion eines Musters als Vorbild einnehmen kann, aber trotzdem Token bleibt und von gleicher Materialität ist wie die sich daran orientierenden Folgetoken. Auf der Ebene der Analyse ist aber ein Muster, das aus musterhaft verwendeten Mehrworteinheiten abstrahiert wurde, tatsächlich ein Type.

Ähnlich verhält es sich mit der ‚Regel‘, die eben nicht die Anwendung derselben ist, sondern einen anderen Status aufweist. Die ‚Schablone‘ ihrerseits ist ebenfalls von anderer Qualität, als das, was sie erzeugt; ebenso das ‚Modell‘: Es ist eine auf „Abstraktion und Idealisierung beruhende (formale) Abbildung wichtiger struktureller und funktionaler Eigenschaften der realen Welt“ und dient dazu, „Vorhersagen über Gesetzmäßigkeiten des Untersuchungsgegenstandes“ (Bußmann 2002, ‚Modell‘) zu machen.

Um auf den Musterbegriff zurück zu kommen: Es ist nicht notwendig, eine abstrakte, kognitive oder tiefensemantische Kategorie ‚Muster‘ zu denken. ‚Musterhaftigkeit‘ lässt sich als Phänomen der Textoberfläche denken, als Phänomen rekurrenten, für bestimmte Kontexte typischen Sprachgebrauchs. Oder, um Wittgensteins Diktum<sup>8</sup> zu bemühen, das Muster liegt – also: definiert sich, entsteht – im Gebrauch.

Das bedeutet nicht, dass in unseren Köpfen nicht auch Vorstellungen über bestimmte Muster existieren können. Doch das ist eine kognitivistische Fragestellung; aus pragmatischer, sprachgebrauchsanalytischer Sicht reicht es vollkommen, Musterhaftigkeit als Phänomen auf der Textoberfläche zu denken – und zu analysieren. Dieser Fokus auf die Textoberfläche würde bei der Verwendung von Begriffen wie ‚Schema‘, ‚Schablone‘, ‚Regel‘ etc. verloren gehen. Für eine Analyse des Sprachgebrauchs, die möglichst auf der Textoberfläche bleiben will, ist deshalb der Musterbegriff passender.

---

<sup>8</sup> „Die Bedeutung eines Wortes ist sein Gebrauch in der Sprache“ (Wittgenstein 1995, 262; §43).

## 2.3 Diskurse

### 2.3.1 Foucault'sche Diskursanalyse

Michel Foucault hat das Forschungsfeld um das zentrale Konzept ‚Diskurs‘ stark beeinflusst. Mit seiner Untersuchung der „Ordnung des Diskurses“ hat er seine Auffassung von ‚Diskurs‘ pointiert formuliert (Foucault 2000; im Rahmen der ‚Inauguralvorlesung am Collège de France‘ vom 2. Dezember 1970), doch bereits in der früheren ‚Archäologie des Wissens‘ ausgebreitet (Foucault 1981). In seiner Inauguralvorlesung findet sich eine der Kernthesen:

Ich setze voraus, dass in jeder Gesellschaft die Produktion des Diskurses zugleich kontrolliert, selektiert, organisiert und kanalisiert wird – und zwar durch gewisse Prozeduren, deren Aufgabe es ist, die Kräfte und die Gefahren des Diskurses zu bändigen, sein unberechenbar Ereignishaftes zu bannen, seine schwere und bedrohliche Materialität zu umgehen. (Foucault 2000, 11)

Solche Prozeduren spannen ein Netz von Aussagen und formen diese zu Diskursen. Sie ermöglichen oder verbieten bestimmte Aussagen und resultieren in einer ‚Episteme‘.

Unter *Episteme* versteht man in der Tat die Gesamtheit der Beziehungen, die in einer gegebenen Zeit die diskursiven Praktiken vereinigen können, durch die die epistemologischen Figuren, Wissenschaften und vielleicht formalisierten Systeme ermöglicht werden. (Foucault 1981, 272f.)

Die Rekonstruktion von Epistemen ist aber nicht die einzige Möglichkeit der Analyse. So möchte Foucault

eine archäologische Analyse konzipieren, die auch die Regelmäßigkeit eines Wissens erscheinen ließe, sich aber nicht vornähme, sie in Richtung der epistemologischen Figuren und Wissenschaften zu analysieren. (Foucault 1981, 274)

Er führt dies am Beispiel der Beschreibung von Sexualität vor. „[A]nstatt das zu beschreiben, was die Menschen über die Sexualität haben denken können“ (also Episteme zu beschreiben), könne man fragen, „ob sich in diesen Verhaltensformen und in diesen Repräsentationen nicht eine ganze diskursive Praxis angelegt findet“ (Foucault 1981, 275). Damit würde eine solche Untersuchung aufzeigen,

wie die Verbote, Ausschlüsse, Grenzen, Aufwertungen, Freizügigkeiten, Grenzüberschreitungen der Sexualität, alle ihre sprachlichen oder nicht-sprachlichen Manifestationen an eine determinierte diskursive Praxis gebunden sind. Sie würde eine gewisse „Sprechweise“ gewiß nicht als letzte Wahrheit über Sexualität, sondern als die eine der Dimensionen, in denen man sie beschreiben kann, erscheinen lassen; und man würde zeigen, wie diese Sprechweise nicht in wissenschaftlichen Diskursen, sondern in einem System von Verboten und Werten angelegt ist. Diese Analyse vollzöge sich

so nicht in der Richtung der Episteme, sondern in der, die man als die der Ethik bezeichnen könnte. (Foucault 1981, 275f.)

In diesen Zitaten sind einige Punkte enthalten, die mir wichtig erscheinen:

1. Der Untersuchungsgegenstand der Diskursanalyse sind nicht nur Episteme, sondern auch Sprechweisen, also Systeme der diskursiven Praxis.
2. Die Prozeduren der Produktion des Diskurses formen also nicht nur Episteme, sozusagen das inhaltlich Sagbare der Diskurse, sondern auch Sprechweisen, die diskursive Praxis.
3. Diese diskursive Praxis ist als regelmäßige Art des Aussagens charakterisiert.

Dieser Teilaspekt des gesamten Forschungsprogramms ‚Diskursanalyse‘ trifft exakt das Interesse der vorliegenden Untersuchung, die nach dem Sprachgebrauch in Diskursen fragt. Die „regelmäßige Art des Aussagens“ scheint dabei an den Musterbegriff anschließbar zu sein, indem die Hypothese formuliert wird, dass musterhafte Sprechweisen zu einer regelmäßigen Art des Aussagens führt.

In seiner ganzen Breite hat sich die Diskursanalyse in vielen geistes- und sozialwissenschaftlichen Disziplinen als fruchtbar erwiesen, obwohl, oder gerade weil Foucault auf exakte Definitionen und die Entwicklung von Methoden verzichtete. Dies hat zur Folge, dass die einzelnen Disziplinen je eigene Vorstellungen von ‚Diskurs‘ und damit verbundene Methoden entwickelten, so auch die Linguistik.

### 2.3.2 Linguistische Diskursanalyse

In der germanistischen Linguistik<sup>9</sup> haben Busse/Teubert (1994) mit dem Terminus ‚linguistische Diskursanalyse‘ als Operationalisierung des Foucault’schen Diskursbegriffs eine Forschungsperspektive vorgeschlagen, die als

Diskurssemantik (die – schon vom Begriff her – nur als diachrone Semantik d. h. als Diskursgeschichte, möglich ist) als eine Erweiterung der Möglichkeit einer linguistisch reflektierten, mit genuin sprachwissenschaftlichen Methoden arbeitenden Wort- und Begriffsgeschichte (Busse/Teubert 1994, 13)

<sup>9</sup> Vgl. für einen konzisen Forschungsüberblick Bluhm u. a. (2000).

charakterisiert wird. Damit wird Diskurs als ein aufgrund semantischer Kriterien zusammengestelltes Korpus an Texten definiert, wobei das Untersuchungsinteresse darin liegt, die diskursiven Beziehungen, „in einem weiten Sinn von Semantik[:] semantische Beziehungen“ (Busse/Teubert 1994, 16) offen zu legen:

Unser sprachwissenschaftliches Interesse an Diskursen entspringt der Absicht, die sprachlichen Manifestationen alternativer Sichtweisen und Vorstellungswelten, Gedanken- und Bedeutungsparadigmen, der epistemischen Voraussetzungen und Leitelemente, die das Thema bzw. den Untersuchungsgegenstand bestimmen, ausfindig zu machen, zu dokumentieren und zueinander in Beziehung zu setzen. (Busse/Teubert 1994, 18)

Das Untersuchungsobjekt ist somit ein definierter Diskurs in Form eines nach semantischen Kriterien zusammengestellten Textkorpus, das

- einen „gewählten Gegenstand, [ein] Thema, Wissenskomplex oder Konzept“ repräsentiert und dessen Texte „untereinander semantische Beziehungen aufweisen und/oder in einem gemeinsamen Aussage-, Kommunikations-, Funktions- oder Zweckzusammenhang stehen“;
- das Korpus ist zudem bezüglich Parametern wie „Zeitraum/Zeitschnitte, Areal, Gesellschaftsausschnitt, Kommunikationsbereich, Texttypik“ eingeschränkt und
- die Texte bilden einen „intertextuellen Zusammenhang“ durch „explizite oder implizite (text- oder kontextsemantisch erschließbare) Verweisungen“;
- das tatsächlich untersuchte Korpus stellt jedoch nur eine relevante Teilmenge des Diskurses dar, da vorab aus forschungspraktischen Gründen eine Beschränkung nötig sei (Busse/Teubert 1994, 14).

Die ausgewählten Texte sollten darüber hinaus repräsentativ „hinsichtlich eines jeweils als Untersuchungsleitfaden gewählten Inhaltsaspekts sein“ und so sei Repräsentativität in der Diskursanalyse „vor allem ein inhaltliches (semantisches) Problem“ (Busse/Teubert 1994, 14f.). Im Vordergrund stünden „inhaltlich begründbare Relevanzkriterien“, deshalb sei es



sinnvoll, beispielsweise Redundanzen zu vermeiden und vornehmlich solche Texte aufzunehmen, die die Struktur und den Verlauf des Diskurses maßgeblich beeinflusst haben; das heißt aber auch, daß die Zusammenstellung des Korpus nicht unabhängig sein kann von einer zuvor erfolgten ersten Inaugenscheinnahme der Texte und einer – schon im Hinblick auf die Untersuchungsziele erfolgenden – Prüfung der Eignung der einzelnen Texte. (Busse/Teubert 1994, 14)

Leicht anders gelagert versucht Warnke (2002) den Foucault'schen Diskursbegriff für die Textlinguistik fruchtbar zu machen. Es geht ihm darum, den Begriff „einer sprachwissenschaftlichen Kompatibilitätsprüfung“ (Warnke 2002, 132) zu unterziehen, was für ihn auch bedeutet, dass vor allem die kulturwissenschaftliche Komponente dieses Konzepts einer kritischen Prüfung unterworfen werden müsse:<sup>10</sup>

[...] die Archäologie M. Foucaults, seine Erklärung der Konstituierung von Epistemen und die Analyse von Machtverhältnissen in der Genealogie sind gerade keine originär linguistischen respektive textlinguistischen Fragestellungen. (Warnke 2002, 132)

Als ausschlaggebend für eine textlinguistische Nutzung des Diskursbegriffs, mit dem Ziel, den Textbegriff zu entgrenzen, sieht Warnke (2002, 133) die Begriffe ‚Ereignis‘ und ‚Serie‘. Der ereignishafte Einzeltext wird damit vor dem „Reproduktionshintergrund betrachtet“, der „immer im Spannungsfeld von Einmaligkeit und wiederholter Vielheit“ steht:

Für die Textlinguistik besagt das Prinzip der Reproduktion, dass die Musterzugehörigkeit eines Textes also im Kern aus seiner diskursiven Einbettung resultiert. Texte sind eben keine singulären Phänomene, sondern sie sind Repräsentanten einer seriell organisierten diskursiven Praxis. (Warnke 2002, 133)

Damit versteht Warnke (2002, 134) Diskurse als Resultat von über Textmuster gesteuerten seriellen singulären Texten, wobei die Diskurse eine regulative Tendenz aufweisen, da sie ein Aussagesystem strukturieren (Warnke 2002, 135f.). „Demnach findet nur *das* textuellen Ausdruck, was im Rahmen der je herrschenden Diskurse erlaubt ist; die Vertextung anderer Aussagen wird zum anarchischen Gegen-diskurs“ (Warnke 2002, 136).

So sieht Warnke (2002, 138) Diskursivität als textkonstitutives Textmerkmal.

<sup>10</sup> Damit bewertet Warnke (2002, 132) auch die Forschung der Critical Discourse Analysis (CDA) als nicht primär (text-)linguistisch, obwohl sie „ohne Zweifel höchst interessante Fragen der Geschichte von Denksystemen, der Soziologie und auch einer Philosophie in Folgezeiten von Ontologie und Epistemologie“ behandle (Warnke 2002, 132).

Der Sinn des entgrenzten Textbegriffs besteht [...] darin, Texte als Teilmen-  
gen größerer kommunikativer Einheiten zu betrachten, die es fraglos gibt  
und die Inhalt und Form von Texten nicht unwesentlich bedingen (Warnke  
2002, 138).

### 2.3.3 Kritik

Die Ansätze von Warnke (2002) und Busse/Teubert (1994) haben unterschiedliche Ziele. Ersterer möchte Diskursivität als textkonstituierendes Merkmal in die Textlinguistik integrieren, letztere möchten den Diskursbegriff für eine linguistische Diskursanalyse fruchtbar machen. Mit Diskursivität als Textmerkmal steht in erster Linie eine Kategorie zur Verfügung, um einen zu untersuchenden Einzeltext hinsichtlich einer neuen Fragestellung, nämlich seiner diskursiven Geformtheit, zu situieren. Busse/Teubert (1994) hingegen nehmen mit ihrer Methode in Anspruch, Diskurse als epistemische Felder zu beschreiben.<sup>11</sup> Die beiden Ansätze widersprechen sich grundsätzlich nicht, obwohl der textlinguistische Begriff der Diskursivität von Warnke (2002) mit dem Aspekt der musterhaften Reproduktion von Texten und dem Fokus auf Textmuster eine leicht andere Perspektive einnimmt, als Busse/Teubert (1994) das tun.

An der Definition von Busse/Teubert (1994, 14) von Diskurs fällt auf, dass Diskurs mit einem Textkorpus gleichgesetzt wird. Auch Busse/Teubert (1994, 15) räumen ein, dass dies nicht im Sinne Foucaults ist: Diskurs ist bei Foucault „nicht in erster Linie ein Textkorpus, sondern sind Beziehungen zwischen einzelnen Aussagen oder Aussageelementen [...] quer durch eine Vielzahl einzelner Textexemplare“. Die Vereinfachung, Diskurs mit einem nach semantischen Kriterien definierten Korpus gleichzustellen, hat entscheidende Auswirkungen: Damit geht die Vielschichtigkeit der Diskurse verloren, die Aussagen formen, und nicht Texte. Die von Foucault postulierte Auflösung der Einheiten von Text und Autor wird damit zumindest in Teilen rückgängig gemacht.

Diese Kritik wird in anderen diskurslinguistischen Arbeiten geteilt<sup>12</sup> und kann in einer alternativen Zusammenstellung des Untersuchungs-

<sup>11</sup> Es ist selbstverständlich, dass auch mit der ‚Diskursivität‘ nach Warnke (2002) letztlich Diskursanalyse betrieben werden kann, doch wäre dies weniger eine originär textlinguistisch, sondern eben diskurslinguistisch ausgerichtete Analyse.

<sup>12</sup> Vor allem Jung (1996) und auch Spitzmüller (2005, 46f.).

korpus und der Wahl anderer Untersuchungseinheiten operationalisiert werden. Dies werde ich weiter unten in Kapitel 5.3 zeigen.

Die Frage nach der Repräsentativität des Korpus zeigt, dass Busse/Teubert (1994, 14f.) diese nach inhaltlichen Kriterien festlegen, was zu einer manuellen Auswahl von „relevanten“ Texten führt. Diese Beschränkung ist nicht nur unnötig, sondern auch kritisch, wenn das Ziel einer Diskursanalyse darin besteht, auch verborgene und unauffällige Strukturen aufzuzeigen. Relevanzempfinden ist ein Ergebnis diskursiver Formationen. Dieses als Kriterium zu wählen, bedeutet, die offensichtlichen Strukturen des Diskurses zu replizieren.

Dass es anders geht, zeigt Wengeler (2003, 294f.), der in sein Korpus aus Zeitungstexten zum Einwanderungsdiskurs alle Texte des Untersuchungszeitraums aufnimmt, in denen Argumentationsmuster vorkommen. Er nimmt keine Unterscheidung in Leittexte oder besonders relevante Texte vor. Es wurden „ausführliche Hintergrundreportagen zum Thema Einwanderung ebenso berücksichtigt wie kurze Berichte über punktuelle Ereignisse oder einzelne Stellungnahmen von Politikern“ (Wengeler 2003, 295).<sup>13</sup> Die linguistische Diskursanalyse fokussiert ihre Untersuchungen auf inhaltlich-semantische Aspekte von Diskursen.<sup>14</sup> Diese Beschränkung liegt wohl darin begründet, dass Busse/Teubert (1994, 13) sie explizit als – zwar erweiterte – Begriffs- und Wortgeschichte sehen. Dies ist eine unnötige Beschränkung, zumal gerade die Linguistik Konzepte und Methoden entwickelte, die den kommunikativen Wert sprachstruktureller und formaler Aspekte betonen. So kritisiert auch Linke (2003a, 40) diese Einschränkung und plädiert für eine Fokussierung auf den Sprachgebrauch:

Sowohl unter begriffsgeschichtlicher als auch unter diskursgeschichtlicher Perspektive geht es in erster Linie um Semantisches bzw. Thematisches, um den Bezug sprachlicher Konkretionen auf mentale Dimensionen, wenn auch mit Blick auf deren soziale Geprägtheit, auf deren Habitualisierung. Die pragmatischen Dimensionen von Sprache, die historischen Dimensionen des *Sprachgebrauchs* werden meist nur insofern beachtet, als sie als konstitutiv für die Modellierung von Bedeutung verstanden werden. Die Frage nach Gebrauchsweisen von Begriffen, den Ausformungen von Texten und den

<sup>13</sup> Für möglichst große Offenheit bei der Zusammenstellung des Untersuchungskorpus sprechen sich auch Bluhm u. a. (2000, 15) aus.

<sup>14</sup> Auch Wengeler (2003) konzentriert sich auf einen inhaltlich klar definierten Diskurs, wobei er eben dem Diskurs zugesteht, dass er sich auch in anderen thematischen Zusammenhängen zeigen kann.

Routinen sprachlichen Handelns steht weitgehend im Dienst inhaltlich-  
semantischer Analysen. (Linke 2003a, 40)

Möchte man diesen inhaltlich-semantischen Fokus verlassen, gilt es, neben dem inhaltlich Sagbaren des Diskurses den Blick für die Sprechweisen in Diskursen zu schärfen. Die Frage lautet demnach weniger, ob Themen, Wissenskomplexe oder Konzepte in intertextuellen Zusammenhängen stehen, sondern vielmehr durch welche Sprachgebräuche diese Zusammenhänge geschaffen werden. Oder mit anderen Worten: Was (an Inhalten) ist typischerweise im Diskurs typischerweise wie (durch welche Sprechweisen) ausgestaltet auffindbar?

Diese Forderung nach dem Einbezug von Sprachgebrauchsmustern in eine Diskursanalyse wirft nun ein anderes Licht auf das Problem der Korpuszusammenstellung, also der Frage nach den Kriterien, die darüber entscheiden, welche Texte Teil des Diskurses sind. Bei der Beschränkung auf inhaltlich-semantische Aspekte eines Diskurses muss aufgrund inhaltlicher Kriterien entschieden werden, ob ein Text Teil des Korpus sein soll; dies „setzt die Kenntnis des Inhalts der in Frage kommenden Texte voraus. So gesehen setzt also schon die Korpusbildung das Verstehen der Texte voraus“ (Busse/Teubert 1994, 16). Sind nun aber bestimmte Sprachgebrauchsmuster das Kriterium für die Bestimmung eines Diskurses, und können diese für eine quantitative Analyse operationalisiert werden, ergeben sich neue Perspektiven der Korpuszusammenstellung. Die Sprachgebrauchsmuster, die als Basis für die Bestimmung des Diskurses dienen, können nach unterschiedlichen Kriterien ausgewählt werden. Sie können z. B. eine formale (wie etwa Modalverbkonstruktionen), aber auch semantisch/lexikalische Ähnlichkeit aufweisen, oder eine Kombination davon. Ein solcher Diskurs wird über so definierte Mengen ähnlicher Sprachgebrauchsmuster bestimmt.

Vielleicht – das muss bei jeder Analyse gesondert entschieden werden – sind in einem späteren Stadium Korrekturen an der Korpuszusammenstellung nach thematischen Kriterien sinnvoll. Klar ist jedoch, dass ein nach dem Kriterium des typischen Sprachgebrauchs definierter Diskurs zu anderen Analysen führt, als thematisch ausgerichtete Diskursanalysen, obwohl natürlich Überlappungen zu erwarten sind.

### 2.3.4 Fazit

Was ist nun mit einer am Sprachgebrauch orientierten linguistischen Diskursanalyse gewonnen? Es ist wichtig festzuhalten, dass eine solche Diskursanalyse nicht die ‚bessere‘ Methode ist. Sie beantwortet aber andere Fragen und tut dies in anderer Art und Weise: korpuslinguistischer und quantitativer. Ich sehe darin die folgenden Vorteile:

1. Das Kriterium des typischen Sprachgebrauchs ist für quantitative Analysen operationalisierbar: In welchen Kontexten welche Präferenzen für Sprachgebräuche liegen, lässt sich statistisch berechnen.
2. Es ist damit möglich, induktiv vorzugehen, sowohl bei der Korpuszusammenstellung als auch bei der Analyse des Korpus. Dies wird der Foucault'schen Vorstellung, dass Diskurse sich gegenseitig überlappen und sich in thematisch heterogenen Kontexten zeigen, eher gerecht.
3. Als Nebeneffekt kann die Einheit des Textes aufgebrochen und mit Aussagen als kleinster Einheit gearbeitet werden. Die Analyse der Sprachgebrauchsmuster in unterschiedlichen Kontexten lenkt die Aufmerksamkeit auf Aussagen, nicht auf Texte. Vielleicht zeigt sich, dass bestimmte Aussagen immer wieder mit ähnlichen Sprachgebrauchsmustern realisiert werden.

Busse/Teubert (1994, 25ff.) pochen mit Recht darauf, dass eine „mit linguistischen Methoden arbeitende Beschäftigung mit Inhalten sprachlicher Zeichen oder Zeichenkomplexe[n]“ nicht aus der Linguistik ausgegrenzt werden dürfe, auch wenn durch ihre Deutungstätigkeit eine vollständige Objektivierung unmöglich sei. Auch eine hier vertretene korpuslinguistische Diskursanalyse, die quantitativ arbeitet und sprachliche Formen stärker berücksichtigt, kommt nicht um subjektive Deutungsakte herum. Doch kann diese die eher auf Inhalte fokussierte Diskursanalyse in wichtigen Teilen ergänzen.

Noch ein Wort zum quantitativen Zugang: Besonders bei der Lektüre von Foucaults „Ordnung des Diskurses“ (Foucault 2000) wird klar, dass die Prozeduren, die Diskurse kontrollieren, selektieren, organisieren und kanalisieren dies außerhalb des Bewusstseins der Sprecherinnen und Sprecher tun. So sieht Foucault (2000, 21f.) das

schreibende Individuum der *Funktion* des Autors unterworfen; was das Individuum

schreibt und was es nicht schreibt, was es entwirft, und sei es nur eine flüchtige Skizze, was es an banalen Äußerungen fallen lässt – dieses ganze differenzierte Spiel ist von der Autoren-Funktion vorgeschrieben, die es von seiner Epoche übernimmt oder die es seinerseits modifiziert. (Foucault 2000, 21f.)

Um doch genau diese Prozeduren lassen sich nur aufzeigen, wenn sie in ihrer Regelmäßigkeit in großen Text- oder Aussagemengen systematisch gefunden werden können – denn sie sind im Einzelfall wahrscheinlich nicht auffallend.<sup>15</sup>

## 2.4 Kultur

Diskursanalyse kann im weiteren Kontext als Konzept der Kulturwissenschaften angesehen werden (Nünning/Nünning 2003, 4). Mit den Kulturwissenschaften wird allerdings ein weites Feld unterschiedlichster Ideen und Methoden ausgebreitet, deren Gemeinsamkeiten nur schwierig zu bestimmen sind. Trotzdem ist es nützlich, den Kulturbegriff aufzunehmen, um die korpuslinguistische Diskursanalyse in diese Richtung weiter zu entwickeln.

Die Sprachwissenschaft hat die Leitvokabel ‚Kultur‘ eher spät entdeckt, während Nachbardisziplinen wie die Geschichte mit der ‚Kulturgeschichte‘ oder die Ethnologie schon länger damit arbeiten. Es scheint deshalb sinnvoll, sich zunächst in diesen Nachbarschaftsdisziplinen nach Verwendungsweisen von ‚Kultur‘ umzusehen.

### 2.4.1 Kultur als Bedeutungsgewebe

Clifford Geertz hat als Ethnologe eine semiotische Auffassung von ‚Kultur‘:

Der Kulturbegriff, den ich vertrete [...] ist wesentlich ein semiotischer. Ich meine mit Max Weber, daß der Mensch ein Wesen ist, das in selbstgespon-

<sup>15</sup> Auch Busse/Teubert (1994, 24) bemerken, dass bei einer diachronen Analyse von Veränderungen des Sprachgebrauchs „nicht jede marginale Veränderung [verdient], notiert zu werden“. Bei einer quantitativen Analyse kann die Entscheidung darüber, ob eine beobachtete Veränderung nun marginal ist oder nicht, durch statistische Tests objektiviert werden. Denn diese erlauben eine differenzierte Bewertung von Frequenzen eines Phänomens in Relation zur Anzahl der gemachten ‚Beobachtungen‘.

nene Bedeutungsgewebe verstrickt ist, wobei ich Kultur als dieses Gewebe ansehe. Ihre Untersuchung ist daher keine experimentelle Wissenschaft, die nach Gesetzen sucht, sondern eine interpretierende, die nach Bedeutungen sucht. Mir geht es um Erläuterungen, um das Deuten gesellschaftlicher Ausdrucksformen, die zunächst rätselhaft scheinen. (Geertz 1987b, 9)

Diese Definition führt Geertz zu dem, was er in Anlehnung an Gilbert Ryle, ‚dichte Beschreibung‘ nennt. Er möchte das menschliche Diskursuniversum erweitern, indem er mit Menschen einer fremden Kultur ins Gespräch kommt, sich mit ihnen austauscht (Geertz 1987b, 20) und so „Zugang zur Gedankenwelt der von uns untersuchten Subjekte erschließt, so daß wir – in einem weiteren Sinn des Wortes – ein Gespräch mit ihnen führen können“ (Geertz 1987b, 35). Dafür sei, wie Geertz ausführt, ein „semiotischer Kulturbegriff ganz besonders geeignet“, mit dem Kultur, verstanden als „ineinandergreifende Systeme auslegbarer Zeichen“ den Kontext, einen Rahmen bildet, in dem Zeichen „verständlich – nämlich dicht – beschreibbar sind“ (Geertz 1987b, 21).

Dabei ist nicht zu vermeiden, dass eine Analyse von Kultur „den Gegenstand selbst prägt – *wir interpretieren zunächst, was unsere Informanten meinen, oder was sie unserer Auffassung nach meinen und systematisieren diese Interpretationen dann*“; ethnologische Schriften sind so also „selbst Interpretationen und obendrein solche zweiter und dritter Ordnung“ (Geertz 1987b, 22/23).

#### 2.4.2 Kultur als semiotisches System

Diese Methode der ‚dichten Beschreibung‘ von Kultur als Bedeutungsgewebe zeigt Parallelen zu Ecos Semiotik auf, „die *alle* Kulturphänomene als Kommunikationsprozesse untersucht“ (Eco 2002, 32). ‚Kulturelle Einheiten‘ bilden die Bausteine von Kulturen, die durch Zeichen repräsentiert und so beobachtbar gemacht werden (Eco 1977, 185f.). Da jedes Zeichen seinerseits wieder durch ein Zeichen definiert werden muss, entsteht eine unendliche Semiose. Daraus ergibt sich ein System von Zeichen, „*das sich aus sich selbst heraus durch aufeinanderfolgende Systeme von Konventionen klärt, die sich gegenseitig erklären*“ (Eco 2002, 77).

Eco zeigt, wie eine so verstandene Kulturanalyse nicht nur, wie im Falle von Geertz, bei fremden Kulturen angewandt werden kann,

sondern auch kulturelle Codes unserer eigenen Kultur semiotisch gelesen werden können. So analysiert Eco beispielsweise Reklame-Botschaften (Eco 2002, 267ff.) oder architektonische Codes (Eco 2002, 325ff.).

Eco ist einer der Vorläufer, die zumindest aus semiotischer Sicht den Kulturbegriff in die Linguistik eingeführt haben. In anderen Bereichen der Disziplin sieht die Verwendung des Begriffs jedoch anders aus: Linke (2003c, 42) beschreibt die Karriere von ‚Kultur‘ in der Sprachwissenschaft und Sprachgeschichte als „beiläufig begonnen“, da sie zunächst bloß die Funktion hatte, den Begriff des ‚Sozialen‘ zu ersetzen. Das Resultat war das ‚Soziokulturelle‘, mit dem

der Kulturbegriff dem „Hohen“ entzogen und zur umfassenden Bezeichnung [wurde] für diejenigen sprachlichen wie außersprachlichen Gewohnheiten, Umgangsweisen, Repräsentations- und Verhaltensformen, die die unterschiedlichen Gruppen einer Sprach- und Kommunikationsgemeinschaft auszeichneten und als deren Bezeichnung der Begriff des Sozialen mit seiner assoziativen Bindung an das Konzept einer Klassengesellschaft nicht mehr adäquat erschien. (Linke 2003c, 43)

‚Kultur‘ ist somit nicht mehr nur der Gegenstand des ‚Hohen‘, sondern auch dann die passende Bezeichnung, wenn Sprache und Sprachgebrauch des täglichen Lebens betrachtet werden.

Doch eine weitere Facette kommt hinzu: Es ist die Einsicht, dass Sprache und Sprachgebrauch nicht nur Ausdruck einer Kultur sind, sondern gestaltend auf sie wirkt (Linke 2003c, 45). Mit Cassirer gesprochen, zurückgehend auf Herder und Humboldt (Linke/Günthner 2006, 11f.), wird die Sprache als „Medium symbolischer Schöpfung und Setzung“ (Linke 2003c, 44) betrachtet.<sup>16</sup> Diese symbolischen Schöpfungen und Setzungen sind letztlich Kultur, weil sie in einer Kommunikationsgemeinschaft durch bestimmte Gewohnheiten hervorgebracht wurden. Aus linguistischer Perspektive ist es nahe liegend, hier bestimmte Formen und Muster des Sprachgebrauchs (Linke 2003c, 45) als konstituierend für Kultur zu verstehen. „Konstituierend für Kultur ist [...] die Emergenz von Gewohnheiten, die ihrerseits als Standardisierung von Verhaltensweisen eines Consoziums zu bestimmen sind und das individuelle Handeln prägen“ (Warnke 2001, 243). Damit kann „Kultur als Gesamtheit der Formen und Muster, die

<sup>16</sup> In Kapitel 4.7 werde ich im Rahmen der Sprachgebrauchsgeschichte nochmals detaillierter darauf eingehen.



zur Deutung von Welt zur Verfügung stehen“ (Warnke 2001, 244), betrachtet werden.

### 2.4.3 Fazit

Der Kulturbegriff führt dort weiter, wo die linguistische Anwendung des Diskursbegriffs Lücken aufweist. Denn während die linguistische Diskursanalyse intertextuelle Bezüge, die hauptsächlich inhaltlich-semantic definiert werden, freizulegen sucht, fußt der Kulturbegriff, wie ihn etwa Linke (2003c) und Warnke (2001) vertreten, auf den eher formal definierten Sprachgebrauchsmustern. Im Rahmen einer Kulturanalyse wird z. B. versucht, kommunikative Praktiken (Warnke 2001, 243) zu beschreiben und damit kulturanalytische Sprachgeschichtsforschung zu betreiben, die von der sprachlichen Verfasstheit von Welt ausgeht (Linke 2003c, 56). Eine solche kulturanalytische Forschung ist an eine Foucault'sche Diskursanalyse anschlussfähig: „Textkultur ist also die Existenz sprachlicher Muster zur Organisation von Aussagen, die kommunikative Absichten verfolgt und damit Teil der pragmatischen Deutungsformen einer Kultur ist“ (Warnke 2001, 245). Die Organisation von Aussagen geht auf eine „Ordnung des Diskurses“ und die Prozesse, die Diskurse steuern, zurück (vgl. Kapitel 2.3). Dass dazu Sprachgebrauchsmuster gehören, ist eine der Kernthesen der linguistischen Kulturanalyse.

### 3 Die Bedeutung musterhaften Sprachgebrauchs

In Kapitel 2 versuchte ich zu zeigen, wie sich eine korpuslinguistische Diskursanalyse von anderen Ansätzen der Diskursanalyse unterscheidet: Sprachgebrauchsmuster bilden dabei die Grundlage, um Diskurse zu identifizieren.

Doch der Zusammenhang zwischen Sprachgebrauch und sprachlicher Musterhaftigkeit, sowie zwischen Sprachgebrauchsmustern und Diskursen ist bislang zu wenig deutlich. Dieses Kapitel soll deshalb diese Zusammenhänge genauer ausleuchten. Im Zentrum stehen die beiden folgenden Fragen:

1. Warum ergeben sich im Sprachgebrauch Muster?
2. Inwiefern repräsentieren Sprachgebrauchsmuster Diskurse?

Die Beantwortung dieser Fragen orientiert sich an folgender These: Musterhafter Sprachgebrauch ist das Resultat von kooperativem sprachlichen Handeln. Bestimmte Sprachgebrauchsmuster zeigen deshalb bestimmte Aspekte sprachlichen Handelns an und können so als charakteristische Eckpfeiler von Diskursen gelesen werden.

#### 3.1 Sprachgebrauchsmuster als Effekte sozialen Sprachhandelns

##### 3.1.1 Formelhafte Textoberfläche – Kontextualisierung

Kontext als Bedingung für Sprechen anzunehmen, ist eines der Grundpostulate der sog. ‚pragmatischen Wende‘. Auch wenn es einerseits fragwürdig ist, dieses Grundpostulat dieser Wende exklusiv zu überlassen,<sup>1</sup> ist es andererseits fruchtbar, die pragmatische Wende als wissenschaftsgeschichtlichen Dreh- und Angelpunkt zu lesen.

---

<sup>1</sup> So verweist auch Feilke (2000, 66) darauf, dass „die Geschichte pragmatisch argumentierender Sprachbegriffe in der Sprachtheorie tiefergehende Fundamente gelegt hat, als in der ‚pragmatischen Wende‘ selbst zum Tragen kommen.“

Allerdings hat sich mit dieser Wende die Blockbildung in der Linguistik – Block 1: die sog. ‚Systemlinguistik‘, Block 2: die sog. ‚Bindestrichlinguistik‘ – verschärft, und in einem zweiten Schritt wurde die Pragmatik als Kind der Wende im systemlinguistischen Sinn umgedeutet und vereinnahmt: „Austin, Wittgenstein, and Grice were hailed as heroes in the 1970s and their insights were quickly integrated into a systemoriented linguistics looking for universal features of language“ (Nerlich 1995, 311). Damit führte die pragmatische Wende zwar zu einer Ausweitung des linguistischen Untersuchungsgegenstandes: Es wurden nun auch Alltagstexte und Gespräche betrachtet. Doch wurde nun parallel zum universalgrammatischen ein universalpragmatisches Erkenntnisinteresse formuliert, das sich auf die Suche nach der Universalität von Sprechakten und deren tiefenstrukturellen Gemeinsamkeiten machte. Feilke (2003, 217) sieht darin aber nur den ersten Teil einer pragmatischen Wende, die sich in den vergangenen 20 Jahren durch eine „zweite pragmatische Wende“ unter geänderten Vorzeichen fortsetzte.

Diese zweite Wende setzt ihre Positionen in einigen Bereichen neu (Feilke 2003, 217ff.):

- Problematisierung der Medialität von Sprache und Text und damit ein neues Interesse mit anderem Blick auf Schriftlichkeit: Die Hinwendung zu Dialog und Diskurs in der ersten pragmatischen Wende führte zu einer Vernachlässigung schriftkonstituierter Sprachlichkeit, die dadurch neu problematisiert wird.
- Kulturalität statt Universalität: Während also die erste pragmatische Wende die Universalität sprachpragmatischer Phänomene betonte, sieht die zweite Wende gerade darin Kulturspezifika.
- Formulierung/Prägung/Lexikon statt Generativität der Kompetenz, Oberfläche statt Tiefenstruktur: Die Fähigkeit, den pragmatischen Wert eines Textes bestimmen zu können, liegt weniger darin, Tiefenstrukturen zu entschlüsseln, sondern über ein *Sprachgebrauchswissen* zu verfügen. Überspitzt formuliert: Die Art der Formulierung entscheidet darüber, ob es sich bei einem Text um einen Wetterbericht oder eine politische Ansprache handelt.

Vgl. dazu die beiden Gruppen von Formeln:

1. „... Durchzug von ... weitgehend trocken ... vereinzelt ...“<sup>2</sup>

2. „... liebe Mitbürgerinnen und Mitbürger ... unser ist ... aber ... gleichzeitig nimmt ... stetig ab ... jeder ... unabhängig von ... ich will ... unser ... muss ... werden ... mehr Wettbewerb ... solidarisch ... ich bin sicher ...“<sup>3</sup>

- Kontextualisierung statt Kontext: Statt den Kontext als Grundlage für das Verstehen zu begreifen, wird Kontext auch als Folge von Text betrachtet: Anscheinend sind bestimmte Elemente in einem Text dafür verantwortlich, dass unterschiedliche Leserinnen und Hörer den gleichen Kontext erzeugen.

Interessant im Zusammenhang der vorliegenden Arbeit sind primär folgende Elemente: Zeichen, mit denen sprachlich gehandelt wird, stellen Kontext her. Damit wird Sprachhandeln konstituierend für Entitäten wie Textsorten oder kommunikative Gattungen. Denn wie Feilke (2003, 212) klar zeigt, sind es z. B. bestimmte Formulierungen, wie jene oben im Wetterbericht, die genügen, um Leser oder Hörerinnen zu veranlassen, jeweils einen ähnlichen Kontext um die Formeln herum aufzubauen. Damit werden aber solche Formeln zu einem entscheidenden Element, die die Textsorte ‚Wetterbericht‘ konstituieren. Notwendig ist demnach eine Art ‚Sprachgebrauchskompetenz‘<sup>4</sup> oder ‚Sprachgebrauchswissen‘, das Wissen um typisierten Sprachgebrauch, das den Sprecherinnen und Hörern erlaubt, „to judge the appropria-

<sup>2</sup> Vgl. Feilke (2003, 211).

<sup>3</sup> Phrasen aus dem „Video-Podcast der Kanzlerin zur Gesundheitsreform“ vom 24. Juni 2006 (Angela Merkel); vgl. <http://www.bundestkanzlerin.de/> (26. 2. 2008).

<sup>4</sup> Mit ‚Kompetenz‘ klingt die Chomskysche Dichotomie ‚Kompetenz vs. Performanz‘ an. Die Sprachgebrauchskompetenz weist mit dieser Kompetenz die Gemeinsamkeit auf, dass sie von den Sprechern der Sprache unbewusst angeeignet wurde. Allerdings sind die Sprachgebrauchskompetenzen der Individuen nicht deshalb ähnlich, weil es ein dahinter stehendes ‚Sprachgebrauchssystem‘ gibt, sondern weil es sich durch das kooperative Sprachhandeln angeglichen hat. Alternativ zur heiklen Bezeichnung könnte ‚Sprachgebrauchswissen‘ verwendet werden.

teness of given utterances in given contexts“ (Fillmore 1976, 90).<sup>6</sup> Ich werde darauf in Kapitel 4.4 zurück kommen.

Mit der Betonung der Kulturspezifität solcher kontextualisierenden syntagmatischen Ausdrücken zeigt sich das Potential, das in der Analyse von Sprachgebrauch liegt: Die Beobachtung der Textoberfläche wird zu einer Möglichkeit, Kultur anders zu fassen, als es z. B. eine historisch-hermeneutische Lesart kann. Auch Linke (2003a, 21) formuliert die Idee, Sprachgeschichte als Kulturgeschichte zu betreiben. Auch darauf werde ich in Kapitel 4.7 zurück kommen.

Mit den Paradigmen der ‚zweiten pragmatischen Wende‘ wird typisierter Sprachgebrauch als Effekt des Sprachhandelns gesehen. Dieser typisierte Sprachgebrauch ist ein Phänomen auf der Textoberfläche. Sprachgebrauchskompetenz ist die nötige individuelle Voraussetzung von Sprachteilnehmerinnen und Sprachteilnehmern für geglücktes Meinen und Verstehen. Es gilt nun im Folgenden noch genauer nach der Sozialität und Kulturalität von Sprachgebrauch zu fragen.

### 3.1.2 Sozialität und Kulturalität von Sprache

Der Sprache wurden in der linguistischen Tradition unterschiedlich starke Verknüpfungen mit Kulturalität zugestanden. So sah Humboldt z. B. eine enge Verknüpfung von Sprachgebrauch, Denken und Kultur, die letztlich so weit geht, dass Sprache den Geist konstituiert.<sup>7</sup>

Zunächst schafft Humboldt (1973, 6) in seiner 1806 erschienenen Schrift „Latium und Hellas“ eine Verbindung zwischen Sprache und „Nation“. Die Sprache sei der wichtigste Faktor, der „Nationaleigentümlichkeit“ ausmache. Nicht „der Wohnort, das Klima, die Religion, die Staatsverfassung, die Sitten und Gebräuche“ machten diese Eigentümlichkeiten aus, es sei allein die Sprache:

5 Deutsch übersetzt als „Welche *sprachliche* Kompetenz ist es [...], die es einem Sprecher ermöglicht, einer gegebenen sprachlichen Äußerung einen *bestimmten* Kontext des Handelns zuzuordnen?“ in Feilke (2000).

6 Fillmore (1976, 88) beschreibt dies als Zuordnung zu „set of worlds in which the discourse could play a role, together with the set of possible worlds compatible with the message content of the discourse“.

7 Humboldt ist nicht der erste Denker in dieser Tradition. Der Topos von der Korrelation von Sprache und Denken findet sich auch bereits im 17. Jahrhundert (z. B. bei Philipp von Zesen, 1651). Vgl. für eine konzise Darstellung dieser Denktradition Stukenbrock (2005, 112ff.).

Ohne sie, als Hilfsmittel zu gebrauchen, wäre jeder Versuch über Nationalitätlichkeit vergeblich, da nur in der Sprache sich der ganze Charakter ausprägt, und zugleich in ihr, als dem allgemeinen Verständigungsvehikel des Volks, die einzelnen Individualitäten zur Sichtbarwerdung des Allgemeinen untergehen. (Humboldt 1973, 6)

Als Erklärung dafür sieht Humboldt den engen Zusammenhang von Sprache und Geist. Sprache sei eben kein „totes Erzeugtes, sondern weit mehr wie eine Erzeugung“ (Humboldt 1973, 34) anzusehen und – letztlich – mit ihrem „Einfluß auf den Entwicklungsgang des Geistes“ diesen konstituierend: „Hierin [in der Spracherzeugung, NB] aber liegt gerade sowohl die Ursach der Verschiedenheit des menschlichen Sprachbaues“ (Humboldt 1973, 35).

Diese enge Verbindung von Geist, Sprache und „Nation“ ist nun aber nicht ohne die Wirkung sozialen Handelns zu denken. Für die Sprache ist „Geselligkeit [...] das unentbehrliche Hilfsmittel zu ihrer Entfaltung“ und sie ist damit „durchaus kein bloßes Verständigungsmittel, sondern der Abdruck des Geistes und der Weltansicht der Redenden“ (Humboldt 1973, 21).

Der Humboldt'schen Denktradition entgegengesetzt steht aber eine Position, die sich an den Leitlinien der Naturwissenschaften orientiert und Sprache im cartesianischen Sinn primär als Mittel zum Zweck, nämlich Gedanken zu übermitteln, modelliert. Sprache wird als ein mentales Modul konstruiert, das dem Menschen angeboren ist und durch kulturelle Faktoren höchstens noch auf die Umwelt hin justiert wird. Heute kann Chomsky als wichtigster Vertreter dieser Denktradition bezeichnet werden.

Jäger (2003, 72ff.) zeigt, wie diese unterschiedlichen Prämissen brisante Auswirkungen auf die Gegenstandskonstitution haben. Ob Sprache als eher kulturell/sozial oder genetisch bedingt – um die Positionen zuzuspitzen – modelliert wird, wirkt sich auf den Modus der wissenschaftlichen Behandlung derselben aus. Diese Modi lassen sich letztlich mit Dilthey „in der Unterscheidung von *verstehenden* Geisteswissenschaften und *erklärenden* Naturwissenschaften auf den Begriff bringen“ (Jäger 2003, 71). So richtig diese Differenz ist, scheint es aber sofort als unsinnig, die Sprache als Objekt der Geistes- oder Naturwissenschaften anzusehen. So schlägt Jäger (2003, 80ff.) auch „Szenarien der dritten Option“ vor, die „der transdisziplinären Natur des Erkenntnisobjektes Sprache“ gerecht werden sollen: Man müsse einräumen,

dass Sprache, Kultur und Geschichte, die das Leben der Gattung seit etwa 50'000 Jahren bestimmen, Teil unseres gattungsgeschichtlich entstandenen biologischen Erbes sind. Die Kulturalität des Menschen ist Teil seines biologischen Programms und seine mentalen Vermögen sind eng mit seiner Leiblichkeit verwoben. (Jäger 2003, 91)

Es ist einleuchtend, eine enge Verknüpfung von Sprache und Geist im Humboldtschen Sinn anzunehmen. Es ist ebenfalls einleuchtend, einen Zusammenhang zwischen Sprache und Kultur zu sehen, wobei es hier auf den ersten Blick wiederum plausibler erscheint, Sprache als kulturell bedingt, statt Kultur als sprachlich bedingt anzunehmen. Doch genau letzteres behauptet Jäger (2005, 51), wenn er vom ‚Eigensinn‘ der Sprache spricht:

Jeder Sinn, der mit den semiologischen Verfahren der Sprache kommuniziert wird, ist seinerseits in semiologischen Verfahren der Sprache generiert worden. Sinnkonstitution ist ein Eigenverfahren der Sprache, keines, das im kognitiven Raum einer mentalen „Sprache des Geistes“ seinen Ort hätte, um von dort in die jeweiligen Einzelsprachen zu gelangen, die dann lediglich als eine Art sozialer Distributor von Mentalesisch fungierten. (Jäger 2005, 51f.)

Durch den Sprachgebrauch, durch „performative Vollzüge“, werden die „semantischen Gehalte von Sprachzeichen“ konstituiert; sprachliche und andere Medien distribuieren so „nicht nur die Inhalte der kulturellen Semantik, sondern sie sind auch wesentlich an ihrer Hervorbringung beteiligt“ (Jäger 2005, 53).

Im Prozess der Kommunikation müssen Sprecherin und Hörer nun das je eigene Bedeutungswissen aktivieren. Dieses Bedeutungswissen ist Sprachgebrauchswissen (oder Sprachgebrauchskompetenz, vgl. Kapitel 3.1.1 auf Seite 43). Wissen darüber, wie Aussagen kontextualisiert werden können, Erwartungen darüber, welche Erwartungen an Kontextualisierungsmöglichkeiten das Gegenüber an die Aussagen hat.

Dieser Prozess ist riskant:

Es gibt keine andere Bürgschaft für das Gelingen der Verständigung als den Umstand, dass die Interaktionspartner den Gebrauch – und das heißt die Semantisierung – der Ausdrucks-Ketten in kulturell vertrauten Sprachspielen interaktiv gelernt haben und dass sich deren hypothetische Intersubjektivität in Kooperationshandlungen und in gemeinsamen Bezugnahmen auf die Welt bewährt hat. Gleichwohl bleibt Kommunikation äußerst fragil, störanfällig und riskant. (Jäger 2005, 55)

Im Zitat weiter oben ist es für Humboldt die ‚Geselligkeit‘, die aus Sprache mehr macht, als ein bloßes Hilfsmittel zur Verständigung.

Erst im sozialen Gefüge ergibt sich das Potenzial für Kultur. Nicht Sprache alleine, sondern Sprachgebrauch als Komponente interaktiver Kommunikation bringt eine kulturelle Semantik hervor. Linke (2008) plädiert deshalb dafür, diese Sozialität von Sprache durch die Betonung ihrer kommunikativen Komponente hervorzuheben:

Das heißt, dass die semiotische Fassung von Kultur zu ergänzen oder richtiger: *einzubetten* ist in eine kommunikative Modellierung von Kultur und dass der Interdependenz dieser beiden Dimensionen von Kultur – der semiotischen und der kommunikativen – vermehrtes Augenmerk gelten sollte. (Linke 2008)

Dabei werden drei Dimensionen von ‚Kommunikation‘ beschrieben, die ‚Vergesellschaftung‘ ermöglichen: Dialogizität, Sozialität und Historizität/Kulturalität von Kommunikation (Linke 2008). Weil Kommunikation dialogisch und sozial funktioniert, ist Sprache, genauer: Sprachgebrauch kulturprägend. Und, andererseits, Kommunikation ist kulturell geprägt, beim Einbezug der historischen Dimension damit auch historisch geprägt.

Sprachgebrauch und Kulturalität bzw. Sozialität stehen in einem wechselseitigen Verhältnis: Sprache ist kulturell bedingt, Kultur ist aber auch sprachlich bedingt. Damit die Wahrscheinlichkeit der Verständigung erhöht werden kann, ist eine Sprachgebrauchskompetenz nötig. Um diese zu entwickeln, sind zwei Elemente Voraussetzung:

1. Sprache muss interaktiv verwendet werden (um mit dem typischen Sprachgebrauch vertraut zu werden).
2. Sprache muss kooperativ verwendet werden (um eine gemeinsame Bezugnahme auf die Welt möglich zu machen).

Beide Elemente klingen bereits bei Humboldt an, wenn er (wie oben erwähnt) „Geselligkeit“ als „unentbehrliches Hülfsmittel“ (Humboldt 1973, 21) für die Entfaltung der Sprache nennt. Und Interaktion und Kooperation führen zu dem, was Feilke (1994) als Sprachgebrauchskompetenz beschreibt. Mit Linke (2008) gesprochen ergibt sich somit eine kommunikative Modellierung von Kultur, mit der die Komponenten Dialogizität (Interaktion) und Sozialität (Kooperation) in die Beziehung zwischen Sprache und Kultur eingewoben werden.



### 3.1.3 Common Sense

Wie definiert sich eine Gemeinschaft? Was sind die Gemeinsamkeiten einer Kultur? Ich bin in Kapitel 2.4.1 der Definition von Geertz (1987b) gefolgt, der Kultur als Bedeutungsgewebe, als Menge von kulturellen Codes beschreibt. Da diese kulturellen Codes durch bestimmte Gewohnheiten hervorgebracht wurden, nämlich durch kooperatives Interagieren, ähneln sie sich und werden von den Mitgliedern der Gemeinschaft als konstitutiv für die Gemeinschaft wahrgenommen.

Man kann den Konsens einer Gemeinschaft darüber, wie kulturelle Codes gelesen und produziert werden, als ‚Common Sense‘ bezeichnen. Damit folgen wir einer Teilbedeutung der philosophischen oder soziologischen Definition von Common Sense, der dort auch oft mit ‚Gemeinsinn‘ beschrieben wird: Es handelt sich um die Art der Auffassung, des Verstehens und der Interpretation von Phänomenen, wie sie die breiten Schichten einer Gesellschaft (Hillmann 1994, ‚Common Sense‘) vornehmen. Sie tun dies nicht aufgrund expliziter Verstandesargumente, sondern nach primären Einsichten des „gesunden Menschenverstandes“ (Metzler 1999, ‚Common sense‘). Der Common Sense fördert dadurch das Zusammengehörigkeitsgefühl der Angehörigen der Gemeinschaft (Hillmann 1994, ‚Gemeinsinn‘).

Feilke (1994, 37ff.) zeigt die Genese des philosophischen Common Sense-Begriffs im Lichte der religiösen Säkularisierung und der Ausdifferenzierung des Wissenschaftssystems. Thomas Reid (1710–1796), wichtigster Vertreter der ‚Schottischen Schule‘, beschreibt Common Sense als „aus der Erfahrung hervorgegangenen Schatz von Basisurteilen (Prinzipien), der [...] allen Menschen gemeinsam ist und vor allem nicht nur kognitive, sondern auch moralische Orientierung bietet“ (Feilke 1994, 40). Das Konzept des Common Sense bietet somit Ersatz für die weggefallene religiöse Erklärungs- und Begründungsmöglichkeit für Alltagswirklichkeit, ohne auf religiöse Dogmen oder wissenschaftlichen Rationalismus zurückgreifen zu müssen. Denn „der Common sense [begründet sich] aber damit, daß es sich gar nicht um etwas Begründungsbedürftiges handelt, sondern um das Leben *in nuce*. Er beruft sich auf die Welt“ (Geertz 1987a, 264).

Auch aus konstruktivistischer und systemtheoretischer Sicht spielt Common Sense-Wissen eine elementare Rolle: Das wahrnehmende Subjekt steht grundsätzlich vor dem Problem der Kontingenz. Seine

Deutung der Welt könnte auch anders sein, seine Selektion einer Art des Handelns könnte auch anders ausfallen. Es gibt keine ‚richtige‘ Deutung, kein ‚richtiges‘ Handeln, sondern nur eine (bezüglich einer Funktion) ‚passende‘ Deutung und ein ‚passendes‘ Handeln.<sup>8</sup> Doch so kontingent und arbiträr diese Selektion ist, so konventionell ist sie. „Indem die Handlungsstruktur in diesem Sinne erst ökologisch – im oikos des Handelns – festgelegt wird, ist die auf diese Art und Weise erworbene individuelle Kompetenz *erst als Folge einer möglichen Performanz feststellbar*“ (Feilke 1994, 63). Der Vollzug der Handlung orientiert sich am Common Sense-Wissen – dadurch wird es sichtbar –, das „als Mittel der Vereindeutigung und Sicherung des Wissens als pragmatisch relevanter Information“ dient. „Der Common sense ist ein auf die ökologischen Bedingungen menschlichen Handelns bezogenes und durch diese Bedingungen pragmatisch konstituiertes und stabilisiertes intuitives Wissen“ (Feilke 1994, 65f.).

Aus systemtheoretischer Sicht kann das Entstehen von Common Sense aus dem autopoietischen Operieren von Systemen erklärt werden. Feilke (1994, 67ff.) zeigt auf, wie sich aus dem Beobachtungsprozess eines Mitglieds eines sozialen Systems eine Struktur ergibt, die Umweltbedingungen definiert, die wiederum den Selektionsprozess beeinflussen.

Ohne zu tief in die Systemtheorie einzutauchen, lässt sich feststellen, dass folgende Aspekte für meine Zwecke interessant sind: Common Sense erscheint uns einerseits in der Form einer Kompetenz, mit deren Hilfe Handeln vereinfacht wird, andererseits als Fundus oder Ablagerung vergangener Selektionsoperationen. Im Modell werden für den Selektionsprozess, der aus den unstrukturierten Informationen Mitteilungen macht (Kneer/Nassehi 1994, 82), Wahrnehmungsmuster und Handlungsschemata hinzugezogen. Ebenso wird versucht, Informationen anschlussfähig an bereits prozessierte Themen zu lesen. Diese Vorgänge beschreiben die Common Sense-Kompetenz.

Gleichzeitig lagert sich aber dieses erfolgreich verwendete Common Sense-Wissen ab. Dies geschieht einerseits durch Rückkopplungsef-

---

<sup>8</sup> Ernst von Glasersfeld prägte im konstruktivistischen Kontext dafür den Begriff ‚Viabilität‘, dessen Namen vom englischen ‚Viability‘ (deutsch: ‚Gangbarkeit‘) abgeleitet ist. Der Terminus meint das Passen im Sinne des Funktionierens. „Das heißt, etwas wird als ‚viabel‘ bezeichnet, solange es nicht mit etwaigen Beschränkungen oder Hindernissen in Konflikt gerät“ (von Glasersfeld 1992, 18).

fekte: Aus den zu Mitteilungen prozessierten Informationen werden Handlungsschemata, Wahrnehmungsmuster und Themenfelder destilliert. Andererseits entsteht im Verbund mit den Operationen der anderen Mitglieder des sozialen Systems die Struktur, die für das soziale System überhaupt konstitutiv ist. Die Struktur – und mit ihr das soziale System – ist das Ergebnis des individuellen Prozessierens von Informationen der beteiligten Subjekte. Diese Struktur bildet nun die Umweltbedingungen, nach denen sich die sich darin operierenden Systeme richten müssen. Man kann diese Struktur als soziale Ordnung bezeichnen. Oder eben als abgelagertes Common Sense-Wissen.

Aus linguistischer Perspektive kann dafür plädiert werden, Sprachgebrauchskompetenz als Teil des Common Sense-Wissens, das Kontingenz durch Konventionalisierung aufhebt, anzusehen. Common Sense im Allgemeinen und Sprachgebrauchswissen im Speziellen sind Ablagerungen vergangener Selektionsoperationen, also vergangener Operationen, die Sinn von Unsinn scheiden.

Der Tatsache der Kontingenz bei der Interpretation von Erfahrungsdaten – ob bei der einfachen Sinneswahrnehmung oder der Wahrnehmung z. B. von Sprache in kommunikativen Prozessen – korrespondiert der Typus des Common sense-Wissens als Mittel der Vereindeutigung und Sicherung des Wissens als pragmatisch relevanter Information. Der Common sense ist ein auf die ökologischen Bedingungen menschlichen Handelns bezogenes und durch diese Bedingungen pragmatisch konstituiertes und stabilisiertes intuitives Wissen. (Feilke 1994, 65f.)

Sprachgebrauchsmuster als Ablagerungen von Selektionsoperationen sind als Phänomene auf der Textoberfläche fassbar. Das ist wichtig für eine Diskursanalyse, die Teile dieses Common Sense-Wissens rekonstruieren möchte; und besonders wichtig für eine korpuslinguistische Diskursanalyse, die quantitativ arbeiten möchte.

### 3.2 Sprache lesen, Kultur lesen

Die Überlegungen zum Zusammenhang von Sprache und Kultur (Kapitel 3.1) haben gezeigt, dass Sprachgebrauchsmuster ein Ausdruck von Konventionalität sind, ohne die sprachliches, und damit soziales Handeln nicht möglich wäre. Sprachgebrauchsmuster lagern sich im Common Sense-Wissen ab.

Dieses Common Sense-Wissen scheint in der Funktion Parallelen mit epistemischem Wissen aufzuweisen, das im Fokus einer Foucault'schen Diskursanalyse steht (vgl. Kapitel 2.3). Wie das epistemische Wissen ist Common Sense-Wissen nötig für soziales Handeln, da es die Kontingenz des Verstehens und die Arbitrarität sprachlicher Zeichen durch Konventionalität für soziales Handeln nutzbar macht. Die Differenzen liegen in den Inhalten dieser Wissensbereiche. Common Sense-Wissen scheint ein Teil des epistemischen Wissens zu sein, da es sehr allgemeine Basisurteile enthält, während epistemisches Wissen darüber hinaus auch sehr detailliertes Wissen umfassen kann.

Trotz dieser Unterschiede wird deutlich, dass über die Sprachgebrauchsmuster ein Zugriff auf epistemisches Wissen erfolgen kann, und damit auf Diskurse. Oder allgemeiner ausgedrückt: Sprachgebrauchsmuster sind Ausdruck sozialen Handelns und bestimmte Sprachgebrauchsmuster können deshalb als Indikatoren für bestimmtes soziales Handeln gelesen werden.

Auch Busse (2005, 35) sieht diesen Zusammenhang von sozialem Handeln und Sprachgebrauch:

Regelmäßigkeiten indes, die in der Ausformung und im Gebrauch sprachlicher Mittel entdeckt werden können, sind nichts anderes als Regelmäßigkeiten in spezifischen Formen sozialen Handelns. Soziales Handeln erfordert immer, konform mit in der jeweiligen gesellschaftlichen Gruppe geltenden Handlungsmustern zu handeln. Dies ist bei sprachlichem Handeln nicht anders als bei anderen Formen sozialen Handelns. Vielleicht ist bei sprachlichem Handeln nur der Konformitätszwang, d.h. die Notwendigkeit der Übereinstimmung jeder Äußerung mit den je geltenden Regelmäßigkeiten, wegen der spezifischen Funktion der Sprache, Kommunikation und koordiniertes Handeln zu ermöglichen, größer als bei anderen Formen gesellschaftlicher Interaktion. (Busse 2005, 35)

So fordert Busse (2005, 39), die Verflechtungen sprachlicher Äußerungsstrukturen (z. B. Sprachgebrauchsmuster) mit dem epistemischen Wissen zu untersuchen und damit den Wandel dieses Wissens und seine Wechselbeziehungen mit gesellschaftlichem Handeln im Rahmen einer Epistemologie zu fassen. Die Linguistik müsse sich hier, sofern sie sozialwissenschaftlich arbeite, als empirische Kulturwissenschaft verstehen (Busse 2005, 42).



## 4 Was sind Sprechweisen?

Während es nun einleuchtend ist, Sprachgebrauchsmuster als Indikatoren für soziales Handeln, und damit für Kultur zu lesen, ist bislang unbeantwortet geblieben, welche Typen von Sprachgebrauchsmustern es gibt und welche Aspekte sozialen Handelns diese anzeigen. In verschiedenen Teilgebieten der Linguistik sind jedoch schon einige Konzepte erarbeitet worden, die Phänomene auf der Textoberfläche zu erfassen versuchen. Dieses Kapitel nennt nun eine Reihe solcher Phänomene, wobei geprüft wird, inwiefern sie als textoberflächlicher Indikator für soziales Handeln, also als Sprachgebrauchsmuster, dienen könnten.

### 4.1 Stil

Im traditionellen soziolinguistischen Verständnis<sup>1</sup> wird Stil als von außersprachlichen Faktoren abhängige Variable betrachtet: Stil wäre also abhängig von Faktoren wie der Sozialgruppenzugehörigkeit der Sprechenden, der Kommunikationssituation, der Redekonstellation, der Gattung der Textproduktion oder abhängig von Erwartungen und Normen (Selting/Hinnenkamp 1989, 2). Damit besteht eine relativ starre Korrelation zwischen solchen Kontexten und Stilen. Ändert sich einer dieser Kontexte, müsste sich auch der Stil ändern; oder anders ausgedrückt: Das, was sich im Sprachgebrauch mit der Veränderung des Kontextes ändert, ist zumindest teilweise Stil.

Diese starre Korrelation, sowie die Annahme, dass außersprachliche Faktoren als unabhängige, also gegebene Variable gesehen werden sollen, wurde verschiedentlich kritisiert.<sup>2</sup> Stattdessen wird postuliert,

---

<sup>1</sup> Z. B. ausgehend von Labov (1972).

<sup>2</sup> Für einen Überblick dient die Einführung von Selting/Hinnenkamp (1989) im Band Hinnenkamp/Selting (1989), der eine Reihe von Arbeiten umfasst, die sich dieser Kritik anschließen und von einem interdependenten Verhältnis zwischen Kontext und Stil ausgehen.

das Verhältnis zwischen Kontext und Stil als interdependent zu betrachten. Somit wird Stilen auch eine Kontextualisierungsleistung zugesprochen:

Stile resultieren also daraus, daß konkretes Verhalten in konkreten Sprachgebrauchssituationen interpretiert wird in Relation zu als solchen relevant gemachten paradigmatischen Alternativen. Insofern impliziert Stilanalyse immer auch *Stilvergleich*. Interaktionsteilnehmer selbst können die jeweils relevanten Alternativen in der Situation selbst durch die Herstellung syntagmatischer und/oder paradigmatischer Kontraste produzieren (Stilwechsel und andere Formen der Stilveränderung, Selbst- versus Fremdstilisierungen *expressis verbis*). In dieser Hinsicht sind Stile also dynamische und in der Situation selbst immer wieder erneut hergestellte und gegebenenfalls modifizierte und auf den Rezipienten zugeschnittene – gleichwohl für diesen rekonstruierbare – Mittel der Signalisierung und Herstellung gemeinsam geteilter, relevanter sozialer und interaktiver Bedeutungen; sie sind Kontextualisierungsmittel, die kraft ihrer interpretativen „Indexe“ auf die jeweils relevanten Interpretationsrahmen verweisen [...]. (Selting/Hinnenkamp 1989, 6)

Stil drückt sich also in der Selektion einer Variante von paradigmatischen Alternativen aus. Durch Stilvergleiche können diese Alternativen sichtbar werden. So schlägt Fix (1991) auch das Verfahren des Vergleichs vor, um Elemente des Stils in Texten zu erfassen. Ebenfalls einer pragmatischen Sichtweise von Stil verpflichtet, definiert sie (und spitzt damit Sandig 1995, 28 zu): „Funktion von Stil wäre also, kurzgefaßt, ein ‚Was‘ durch ein ‚Wie‘ im Hinblick auf ein ‚Wozu‘ auszudrücken“ (Fix 1996, 310). Diese Kombination von Wie und Wozu zeigt sich im Unikalen des Textes, jedoch erst im Vergleich mit dem Überindividuellen. Jeder Text ist unikal, verfügt jedoch über einen Stil, der sich an überindividuellen Erfahrungen orientiert. Durch den Vergleich des Textes mit einem „individuellen Konstrukt“ (Fix 1991, 142), das sich aus der Lektüreerfahrung ergibt (Fix 1991, 145), kann das Überindividuelle des Textes erfasst werden.

Mit der linguistischen Erforschung von Stil wird zweierlei versucht: Erstens wird nach den Funktionen von Stil und dem Zusammenhang von Kontext und Stil gesucht. Im Prinzip muss man hier zu ähnlichen Schlüssen kommen, wie wenn man nach den Gründen für die Entstehung von musterhaftem Sprachgebrauch durch soziales Handeln fragt. Zweitens ist ein Ziel der Stilforschung, die Elemente auf der Textoberfläche zu benennen, die Stil ausmachen. Interessant für meine Fragestellung sind die Überlegungen zu Sprachgebrauch, dem

stilbildende Funktion nachgewiesen werden kann. Solcher Sprachgebrauch muss einerseits typisch für eine bestimmte Textfunktion oder eine Textsorte sein, andererseits Stilfunktion aufweisen. Diese Stilfunktion, z. B. wie oben dargestellt definiert als Ausdruck des „Was‘ durch ein ‚Wie‘ im Hinblick auf ein ‚Wozu‘“ (Fix 1996, 310) oder als „Kontextualisierungsleistung“ (Selting/Hinnenkamp 1989, 6), kann korpuslinguistisch nur schwer operationalisiert werden. Bei einer korpuslinguistischen Analyse von musterhaftem Sprachgebrauch, der z. B. typisch für bestimmte Textsorten ist, wird ein Teil der gefundenen Muster Stilfunktion haben. Wie groß dieser Teil ist, d. h. wie zuverlässig die statistische Methode der Erkennung von typischen Sprachgebrauchsmustern für eine Erkennung von Stilmerkmalen arbeitet, wird sich im Experiment erweisen.

## 4.2 Kommunikative Gattungen

Der von Luckmann (1986)<sup>3</sup> konzipierte und von Günthner/Knoblach (1994) linguistisch verfeinerte Begriff der ‚kommunikativen Gattung‘ beschreibt ein konventionalisiertes kommunikatives Verfahren der Gesellschaft:

Diejenigen kommunikativen Vorgänge, die typisch wiederkehrend und deren regelmäßige Bewältigung von gesellschaftlicher Relevanz ist, bilden typische Muster aus, an denen sich Handelnde orientieren können. Kommunikative Gattungen bezeichnen diejenigen kommunikativen Prozesse, die sich gesellschaftlich verfestigt haben. (Günthner/Knoblach 1994, 695f.)

Damit erleichtern kommunikative Gattungen gesellschaftliches Handeln, da sie für bestimmte Akteure, Situationen und Funktionen ein Muster anbieten, nach dem kommunikativ gehandelt werden kann. Erleichternd kommt dazu, dass die Erwartungen, kommunikativen Gattungen zu folgen, wechselseitig sind: Ego folgt einer Gattung und erwartet, dass Alter dies auch tut und Ego weiß, dass Alter ebenfalls erwartet, dass Ego das tut.<sup>4</sup>

<sup>3</sup> Vgl. für einen Kurzüberblick auch Nünning/Nünning (2003, 218).

<sup>4</sup> Diese Erwartungserwartungen zeigen, dass das Wissen um kommunikative Gattungen bei den Sprachbenutzern vorhanden ist. Sie resultieren nämlich in sog. ‚Ethnokategorien‘, in im Sprachgebrauch ersichtlichen Bezeichnungen für kommunikative Gattungen, wie z. B. Textklassen, aber auch in weniger festen Formen wie der Gattung ‚Vorwurf‘. Dieses Wissen konstituiert die Klassen durch ihre Anwendung mit (Günthner/Knoblach 1994, 704).



Es sind diese Erwartungserwartungen, die kommunikative Formen zu Gattungen verfestigen (Günthner/Knoblauch 1994, 702). Als Formalisierung beschreiben Günthner/Knoblauch (1994, 703)

die Kombination verschiedener verfestigter (rekurrenter) Elemente sowohl auf der paradigmatischen als auch syntagmatischen Ebene [...], d. h. [die Formalisierung] umfasst Verfestigungen im Bereich der Selektion von Elementen (beispielsweise bestimmte lexikalische Verfestigungen) wie auch im Bereich der Kombination dieser Elemente (beispielsweise bestimmte Abfolgemuster und Handlungsschritte). (Günthner/Knoblauch 1994, 703)

Damit sind kommunikative Gattungen komplexe Formen von einfacheren Sprachgebrauchsmustern wie typischen Wortkombinationen oder Phraseologismen. So würden Sprichwörter oder ein Gruß/Gegengruß-Paar nur als kleine Formen kommunikativer Gattungen bewertet, da sie zwar mehr oder weniger verfestigt sind, jedoch keine hohe Komplexität und keine ausgeprägte Ablaufform aufweisen.<sup>5</sup> Günthner/Knoblauch (1994, 703) nennen solche schwächer verfestigten und/oder formalisierten kommunikativen Formen „kommunikative Muster“. Auch „Entrüstungsgeschichten, Streitgespräche, Belehrungen oder Beratungssendungen im Rundfunk“, die „durchaus strukturiert [sind] und [...] gattungsähnliche Züge auf[weisen]“, jedoch weniger komplex formalisiert und weniger verpflichtend festgelegt sind, fallen darunter.

Kommunikative Gattungen sind nicht einseitig von sozialen Kontexten bestimmt, sondern konstituieren diese Kontexte selber auch. So konstituieren kommunikative Gattungen z. B. institutionelle Bereiche wie das Rechtssystem, und darin die Produktion und Plausibilisierung von Erzählungen, oder politische Gruppierungen (Günthner/Knoblauch 1994, 701, 712):

Kommunikative Muster und Gattungen sind gleichsam die Institutionen innerhalb der Kommunikation. Die kommunikativen Probleme, die von einer gewissen Relevanz für die Handelnden sind, finden ihren Niederschlag in festgelegten Formen, die bis zu gattungsartigen Verfestigungen reichen. Diese Verfestigungen beschränkt sich keineswegs auf althergebrachte Formen der traditionellen mündlichen Kultur, wie sie von der Volkskunde bislang betrachtet wurde. Wie der Überblick über die verschiedensten Gattungen der institutionalisierten Kommunikation zeigt, neigen auch moderne und

<sup>5</sup> Linke (2009) schlägt für diese ‚kleinen Formen‘ die Bezeichnung ‚kommunikative Praktik‘ vor, behält aber daneben ‚kommunikative Gattung‘ bei. Anders Fiehler u. a. (2004, 99), die sich dafür aussprechen, nur den Terminus ‚kommunikativen Praktiken‘ zu verwenden.

„postmoderne“ Formen zur „sekundären Traditionalisierung“. Gattungsmuster bilden sich nicht nur für Mitteilungen auf Electronic Mail aus, sondern auch in „anspruchsvollen“ Werbespots, in Fernsehdiskussionen oder in der Unternehmensberatung. Kommunikative Muster und Gattungen sind somit auch Antworten auf die Anforderungen der modernen Kommunikationskultur und bilden damit ein Verbindungselement zwischen dem subjektiven Wissensvorrat und den sozialen Strukturen einer Gesellschaft. (Günthner/Knoblauch 1994, 715f.)

Für die analytische Erfassung kommunikativer Gattungen sind die drei Strukturebenen wichtig, auf denen sich die Gattungen anhand bestimmter Merkmale festmachen lassen (Günthner/Knoblauch 1994, 704ff.):

1. Binnenstruktur: Prosodische, lexiko-semantische, morpho-syntaktische, rhetorische, mimisch-gestische Merkmale, Phraselogismen, Formeln, Gliederungsmerkmale und Superstrukturen, inhaltliche Verfestigungen (Themenbereiche) etc.
2. Situative Realisierungsebene: Analyseebene für interaktive kommunikative Gattungen, auf der zwar auch die Elemente der Binnenstruktur gelten, darüber hinaus aber für interaktive Formen spezifische Phänomene beobachtet werden können (Ritualisierungen, Aushandlungsprozesse etc.).
3. Außenstruktur: Definitionen wechselseitiger Beziehungen, kommunikativer Milieus und kommunikativer Situationen sowie der Auswahl von Akteurstypen.

Interessant an den kommunikativen Gattungen sind ihre theoretische Begründung im sozialen Handeln. In einer späteren Arbeit betont Günthner (2003, 193) diesen Aspekt erneut:

Durch Kommunikationsvorgänge werden also kulturelle Vorstellungen, Konventionen, Werte etc. konstruiert und reproduziert, welche wiederum zugleich die Art, wie wir sprechen und handeln bzw. wie wir die Äußerungen und Handlungen des Gegenüber [sic!] interpretieren, beeinflussen. (Günthner 2003, 193)

Allerdings liegt das Erkenntnisinteresse hier vor allem bei der „kontextbezogenen, lebensweltlich verankerten und dialogischen Verwendung“ von Sprache (Günthner 2003, 204), also vornehmlich (zumindest konzeptionell) mündlicher (Alltags-)Sprache.<sup>6</sup> Diese Einschrän-

<sup>6</sup> Vgl. auch Günthner/Knoblauch (1995) und Günthner (1995).

kung des Erkenntnisinteresses schmälert aber die Nützlichkeit des Konzepts nicht. Denn es spricht nichts dagegen, dass kommunikative Gattungen nicht auch in schriftlicher Sprache, z. B. Zeitungstexten, wie ich sie untersuche, gefunden werden können. Prinzipiell können auch kommunikative Gattungen als Teil von Common Sense-Wissen (vgl. Kapitel 3.1.3) angesehen werden, wobei sie eher prozesshaftes, kommunikatives Wissen repräsentieren.

Für eine quantitative korpuslinguistische Studie sind kommunikative Gattungen dann nützlich, wenn sie als Sets von Merkmalen der Textoberfläche operationalisiert werden können. Damit sind vor allem Strukturmerkmale der Binnenstruktur mögliche Indikatoren; teilweise können Metainformationen für die Beschreibung der Außenstruktur hinzugezogen werden, z. B. wenn die Textsorte oder der Autor/die Autorin eines Textes und dessen/deren Funktion in die Analyse einbezogen werden kann.

Bei der induktiven Analyse von Korpusdaten könnte musterhafter Sprachgebrauch Hinweise für kommunikative Gattungen, oder wenigstens kommunikative Muster, geben. Günthner/Knoblach (1994, 703) scheinen nur komplexe Muster als kommunikative Gattungen im engeren Sinn zu akzeptieren; diese durch quantitative, induktive korpuslinguistische Analysen freizulegen, dürfte schwierig sein. Berücksichtigt man das ganze Kontinuum, das mit den beiden Achsen Verfestigung und Formalisierung zwischen schwach und stark aufgetan wird, können mit einer korpuslinguistischen Analyse jedoch über musterhaften Sprachgebrauch Beispiele für kommunikative Gattungen und Muster gefunden werden.

### 4.3 Mentalitäten und diskurssemantische Grundfiguren

Der von Hermanns (1995) für die Linguistik fruchtbar gemachte Begriff der ‚Mentalität‘ beschreibt ein kollektives Fühlen, Denken und Wollen, das die Grenzen des Sagbaren in einer bestimmten Gesellschaft zu einer bestimmten Zeit absteckt. Gemäß Hermanns (1995, 79ff.) lässt sich Mentalitätsgeschichte als Begriffsgeschichte verstehen und der Mentalitätsbegriff ist damit an das Konzept der linguistischen Diskursanalyse (vgl. Kapitel 2.3) anschließbar.

Hermanns (1995, 90) plädiert dafür, mit der Mentalitätsgeschichte „die Zusammenhänge wieder in den Blick zu rücken, denen Texte ihre Existenz und ihre Form und ihren Sinn verdanken“ und wehrt sich gegen korpuslinguistische Zugänge, denen er unterstellt:

Sind die Texte einmal im Computer, braucht man diese Texte daher gar nicht mehr zu lesen, so das Ideal der Korpuslinguistik. Als die Teile eines Korpus warten Texte sozusagen nur noch darauf, dass man sie verzettelt. Die Verzettelung der aus dem Text gewonnenen Belege ist die zweite Phase dessen, was man eine Textzerstörung nennen könnte. Dessen erste Phase aber ist bereits die Sinnlosmachung eines Textes dadurch, dass man ihn aus seinem intertextuellen und historischen Zusammenhängen isoliert. (Hermanns 1995, 90)

Der einzelne Text soll also wieder in den Vordergrund der Analyse treten. Trotzdem bleibt die Frage der Operationalisierbarkeit von Mentalitäten unklar, was auch Scharloth (2005a, 134) kritisiert und deshalb die Analysekategorie der diskurssemantischen Grundfigur von Busse (1997, 20) aufgreift und sie in semantischer Perspektive genauer definiert, um sie als Analysekategorie operationalisierbar zu machen. Denn mit dieser Kategorie ließe sich Mentalität besser fassen, da sie als diskursübergreifendes Phänomen „grundlegende Elemente der Organisation des umfassenden semantischen Systems“ sind (Scharloth 2005a, 140) und zur „epistemisch-kognitiven Grundausrüstung der Produzenten“ gehören (Scharloth 2005a, 137).<sup>7</sup>

Scharloth (2005a, 140) bietet zwei semantische Kategorisierungen an, um diskurssemantische Grundfiguren zu fassen:

Zum einen können sie als semantische Oppositionen konzeptualisiert werden, die eine große Anzahl von semantischen Feldern strukturieren, zum anderen können sie als grundlegende Merkmalsoppositionen beschrieben werden, die semantische Felder von einander abgrenzen. (Scharloth 2005a, 140)

So ist in unserer Kultur beispielsweise die Merkmalsopposition ‚belebt‘ vs. ‚unbelebt‘ aktiv, um physische Gegenstände einzuteilen; als semantische Opposition, die viele semantische Felder strukturiert, rekonstruiert Scharloth (2005a, 140ff.) die Dichotomie ‚Natur‘ vs. ‚Künstelei‘ für das letzte Drittel des 18. Jahrhunderts.<sup>8</sup> Diese Opposition strukturiert nicht nur semantische Felder in Texten, sondern z. B. auch in zeitgenössischen Bildern, die Alltagsszenen ‚natürlicher‘ und ‚affectirter‘ Manier zeigen (Scharloth 2005a, 142). So erlaubt die Analysekategorie der diskurssemantischen Grundfigur eine umfassende

<sup>7</sup> Vgl. auch Busse (1997, 19).

<sup>8</sup> Weitere diskurssemantische Grundfiguren finden sich in Scharloth (2005b).

Beschreibung kulturhistorischer Prozesse, in der „Gesellschaftskritik [...] tiefensemantisch vorstrukturiert“ ist und damit „auf einem Feld des Sagbaren statt[findet], das (unter anderem) von der Opposition ‚Natur‘ vs. ‚Künstelei‘ abgesteckt wird“ (Scharloth 2005a, 144).

So überzeugend mit dem Konzept der diskurssemantischen Grundfigur Räume des Sagbaren abgesteckt werden können, so schwierig ist eine korpuslinguistische Nutzung des Konzepts, ist es doch explizit auf einer tiefensemantischen Ebene angesiedelt (Scharloth 2005a, 140). Der Schlüssel für einen korpuslinguistischen Zugang müssten Phänomene auf der Textoberfläche sein, die als Indikatoren für diskurssemantische Grundfiguren gelesen werden könnten. Doch ähnlich wie bei den kommunikativen Gattungen (Kapitel 4.2) müsste es bei einem induktiven Zugang zu Korpusdaten doch möglich sein, Hinweise für diskurssemantische Grundfiguren zu finden, deren Plausibilität anschließend inhaltsanalytisch geprüft werden könnte. Ein solcher Zugang hätte – um der kritischen Bewertung der Korpuslinguistik von Hermanns zu widersprechen – gerade den Vorteil, nicht von einzelnen Texten auszugehen, sondern von größeren Datenmengen, in denen nach textübergreifenden Sprachgebrauchsmustern gesucht werden kann. Damit könnte die mögliche Wirkung unterschiedlicher Grundfiguren über mehrere Texte hinweg, eben in Diskursen, berücksichtigt werden.

Interessant ist, dass es mit den diskurssemantischen Grundfiguren gelingt, die linguistische Diskursanalyse zu einer „Methode der Analyse von Kulturen“ (Scharloth 2005a, 147) zu entwickeln. Dies dadurch, dass es mit den diskurssemantischen Grundfiguren möglich wird, sich von der Analyse sprachlicher Zeichen zu lösen und alle Zeichensysteme mit einzubeziehen. Hier könnte die (Text-)Korpuslinguistik zumindest Evidenzen im System sprachlicher Zeichen liefern, die in eine umfassendere Analyse integriert werden können.

#### 4.4 Typik des Verhaltens und idiomatische Prägungen

Feilke (1993) stellt die idiomatische Prägung ins Zentrum seiner theoretischen Überlegungen, die „Sprache als soziale Gestalt“ (Feilke 1996) auffassen. Ausgangspunkt ist jedoch die Modellierung einer Com-

mon Sense-Kompetenz als Kontextualisierungskompetenz, die „eine wichtige Grundlage unserer Fähigkeit [ist], gemeinsame Kontexte für Meinen und Verstehen zu *erzeugen*“ (Feilke 1993, 366). Denn Kommunikation ist grundsätzlich kontingent; um einen sinnvollen Verstehensprozess auszulösen, muss Kommunikation durch Common Sense-Wissen strukturiert und damit anschlussfähig gemacht werden.

Idiomatische Prägungen sind das Resultat von Konventionalisierungen von Interpretationen, die ausdrucksseitig diese Konventionalisierung widerspiegeln. „So können über die Prägung bestimmter *Ausdrücke* Ressourcen des Vorverständnisens für die Kommunikation geschaffen und gesichert werden“ (Feilke 1993, 367). Diese „Typik des Gebrauchs“ (Feilke 2003, 209) ermöglicht überhaupt erst pragmatisch brauchbare Sprache.

Damit wird die Beziehung zwischen sozialem Handeln und der Typik des Sprechens deutlich: „Soziales Handeln wird erst als solches erkennbar durch eine Typik des Verhaltens“ (Feilke 1993, 368). Diese Typik wird auch sprachlich realisiert. Eine Typik des Sprechens macht soziales Handeln möglich. Da aber das ‚genau so‘-Sprechen nicht vollständig durch entsprechend geformte Intentionen erklärt werden kann, muss dieses ‚genau so‘ durch eine „idiomatisch bestimmte Ausdrucks-Kompetenz“ (Feilke 1993, 369) modelliert werden. „Ausdrücke können in diesem Sinne gewissermaßen als Handlungsmodelle fungieren, indem sie Schemata sozialer Koorientierung indizieren“ (Feilke 1993, 369). Somit ist auf der Textoberfläche in Gestalt der idiomatischen Prägungen sichtbar, was als Typik auf der Ebene sozialen Handelns erscheint.

Feilke (1996, 272) führt eine Systematik von unterschiedlichen Prägungen auf; er unterscheidet soziale, textuelle und Topik-Prägungen, die weiter differenziert werden. Zu sozialen Prägungen gehören Formeln wie *Das macht 3,50!* oder *Ich schwöre...*, als textuelle Prägungen werden z. B. *Kennst du schon den?* oder *Unter uns gesagt...* kategorisiert und Topik-Prägungen sind beispielsweise *Reisen bildet* oder *Wo Rauch ist, ist auch Feuer*.

Routinen des Sprachgebrauchs können zu eigentlichen Textroutinen erweitert werden. Das Wissen um solche Textroutinen ermöglicht den Sprecherinnen und Sprechern, Sprachgebrauch zu kontextualisieren. Feilke (2003, 210f.) zeigt das anhand eines Experiments, in dem „in

Text 1	Text 2	Text 3
... Durchzug von ...	... auf den Trümmern	... bisher kaum wahrge-
... weitgehend trocken ...	errichtet ...	nommen ...
... vereinzelt ...	... Sorgen und Mühen ...	... werden herausgestellt
	... am Markt bestehen ...	...
	... vorbildliche Zusam-	... Anhand sogenannter
	menarbeit ...	...
	... Die Konkurrenz	... werden kontrastiert ...
	schläft nicht ...	... es zeigt sich ...
	... das Tanzbein schwin-	
	gen ...	
	... erhebe mein Glas ...	

**Tabelle 4.1:** Experiment ‚Textroutinen‘ von Feilke (2003, 211), bei dem Studierende Hypothesen über mögliche Kontexte nennen müssen, in denen diese Ausdrücke zu erwarten sind.

mehreren Seminaren Studierenden verschiedener Studienrichtungen“ eine Reihe von Ausdrücken, die drei Texten entnommen worden waren, vorgelegt wurden; Aufgabe war, mögliche Kontexte dazu zu nennen (vgl. Tabelle 4.1).

Dabei zeigte sich, dass die muttersprachlichen Studierenden weitgehend übereinstimmende Einschätzungen zu den Kontexten nannten; Schwierigkeiten hatten Studierende, die zwar über gute Deutschkenntnisse verfügten, jedoch nicht Deutsch muttersprachlich waren, und, bei Text 3, Studierende der ersten Semester.<sup>9</sup> Dies zeigt, dass nur wenige Ausdrücke reichen, um diese bezüglich Thema, aber auch Textsorte, illokutivem Aufbau und situativem Kontext einordnen zu können. Doch noch erstaunlicher ist, dass dies funktioniert, obwohl

dem zugrunde liegenden Ausdrucksmaterial [...] beinahe alles [fehlt], was nach den Lehrbüchern der Textlinguistik die linguistische Substanz von Texten ausmacht. Das ist meines Erachtens ein Erklärungsproblem

(Feilke 2003, 212). So können semantische Integrationsmittel der Textbildung, wie Implikationen, Präsuppositionen, Folgerungen, konversationelle Implikaturen etc., ihre Wirkung kaum entfalten, da die Ausdrücke gar keinen Satzcharakter aufweisen. Es gibt keine lexikalischen oder grammatischen Bezüge und kaum Hinweise auf Textthemen. In Beispiel 2 kann der illokutive Textaufbau rekonstruiert werden, ohne

<sup>9</sup> Die Herkunft der Ausdrücke soll nicht verschwiegen werden (Feilke 2003, 228f.): Text 1: Wetterbericht; Text 2: Aus einer Sammlung von Musterreden „Betriebsjubiläum – der Inhaber spricht“; Text 3: Abstract eines wissenschaftlichen Aufsatzes.

dass Sätze als Sprechakte analysiert werden könnten; Handlungen müssen also „bereits unterhalb der Satzebene illokutiv spezifizierbar sein“ (Feilke 2003, 212). Und:

Zu keinem der Beispiele standen irgend welche Kontextinformationen zur Verfügung. Ja, man kann sogar davon ausgehen, dass der komplexe Kontext zu Beispiel 2, den die Studierenden ohne Probleme konstruieren, ihnen ausgesprochen fern liegt. Auch die Abwesenheit von Kontext ist texttheoretisch eigentlich nicht vorgesehen. (Feilke 2003, 212)

Das Experiment plausibilisiert die oben referierten Überlegungen zur Rolle der idiomatischen Prägungen, die, in semasiologischer Perspektive, als Sprachmaterial übereinstimmende Vorstellungen typischen sozialen Handelns ermöglichen. Im Experiment wurden eben nicht beliebige Textfragmente verwendet, sondern syntagmatische Ausdrücke, die den „konnotativen Mehrwert“ (Feilke 2003, 213) bieten, der sich durch die Konventionalisierung ergeben hat. Aus phrasologischer Perspektive bemerkenswert ist die Tatsache, dass es sich dabei nicht ausschließlich um Ausdrücke hoher Idiomatizität handelt, sondern hauptsächlich um Kollokationen oder Routineformeln (Burger 1998, 50/52) mit „gänzlich regulärer Struktur und kompositioneller Bedeutung“, die aber „hochgradig konventionell und zeichenhaft“ sind (Feilke 2003, 213).<sup>10</sup> So wird das breite Bedeutungsspektrum von *Durchzug* im ersten Textbeispiel (vgl. Tabelle 4.1 auf der vorherigen Seite) durch die Erweiterung zu *Durchzug von* bereits enorm reduziert, hauptsächlich auf den Kontext Wetter.<sup>11</sup> Allerdings findet Feilke (2003, 214) im DeReKo IDS (o. J.)-Korpus bloß 11 Belege für *Durchzug von*, was zeigt, dass nicht absolute Frequenzen rekurrenten Auftretens eines Ausdrucks dessen Typik bestimmen, sondern zu Textsorten relative Frequenzen. Das ist – um auf Teil II der vorliegenden Arbeit vorzugreifen – eines der Kernprobleme korpuslinguistischer Analysen, das jedoch inzwischen mit der Hilfe statistischer Methoden gut gelöst werden kann.

<sup>10</sup> Interessant ist, dass selbst ein unscheinbarer Ausdruck wie *in der Sonne sitzen* idiomatisch ist: Eine strukturell analoge Formulierung *im Mond sitzen* ist ungebräuchlich; es handelt sich also um Sprachgebrauchswissen, das *unter der Sonne sitzen* o. ä. eher nicht zu sagen zulässt (Feilke 1996, 126).

<sup>11</sup> Neben einer Korpusrecherche, z. B. im DeReKo IDS (o. J.)-Korpus, wird dies auch mittels entsprechender Suchanfragen in einer Suchmaschine wie „Google“ deutlich: Die Belegmenge von *Durchzug* wird durch *Durchzug von* um etwa den Faktor 30 verkleinert, wobei die Mehrzahl der Treffer im Kontext von Wetter und Militär (*Durchzug von Militär*) angesiedelt zu sein scheinen (Anfrage auf [www.google.com](http://www.google.com) vom 6. Juli 2007).



Die Kombination mehrerer idiomatisch geprägter Ausdrücke verringern die möglichen Kontextualisierungen weiter, wobei unter solchen Ausdrücken nicht zwingend Mehrwortausdrücke zu verstehen sind, wie *vereinzelt* in Text 1 zeigt. An diesem Wort zeigt sich auch, dass eine bestimmte morphologische Ausprägung (also das Partizip Perfekt von *vereinzeln*) hinsichtlich der Textsorte Wetterbericht idiomatisch geprägt ist. Die Form *vereinzeln* wird, was eine Recherche im DeReKo IDS (o. J.)-Korpus zeigt, kaum verwendet.<sup>12</sup>

Von den bisher behandelten Ansätzen, die versuchen, die Beziehung zwischen sozialem Handeln und Sprachgebrauch zu definieren, gehören die oben referierten Thesen für einen korpuslinguistischen Ansatz zu den vielversprechendsten Ansätzen. Die Thesen ergeben sich aus einem neuen Verständnis von Pragmatik:

Statt nach dem Einfluss des Kontextes auf das Verstehen des Textes zu fragen, wird die Fragerichtung umgekehrt: Welchen Kontext erzeugt ein gegebener Text? Und welche Elemente eines Textes sind dafür verantwortlich, dass Leser bzw. Hörer gleichsinnig einen bestimmten Kontext erzeugen? Wie funktioniert die Kontextualisierung der Texte? Welche sprachlichen Eigenschaften der Texte sind dafür verantwortlich? (Feilke 2003, 219)

Diese Umkehrung der Fragerichtung und die Erkenntnis, dass letztlich Sprachgebrauchsmuster (oder eben: typische Ausdrücke, idiomatische Prägungen) auf der Textoberfläche die Kontextualisierung ermöglichen, sind die Schlüssel für einen korpuslinguistischen, induktiven Zugang zu Sprachgebrauchsanalysen, die sich als Diskursanalyse verstehen. Wenn die Sprachgebrauchsmuster erkannt werden können, sind Rückschlüsse ihrer typischen Kontextualisierungsleistungen und die Verortung dieser Muster in Diskursen möglich. Dazu ist es nötig, die Distribution dieser Muster in einem Korpus, deren Veränderungen bezüglich Frequenzen, Form oder bezüglich deren Vorkommen mit anderen Mustern zu beobachten.

Das Problem wird also darin liegen, eine korpuslinguistische Methode zu entwickeln, die solche geprägten Ausdrücke auffindet. Dabei ist die Erkenntnis wichtig, dass solche Ausdrücke immer nur prägend sind in Bezug auf eine Textfunktion, z. B. eine Textsorte. Dies

<sup>12</sup> Auch hier wird eine zentrale Fragestellung der korpuslinguistischen Annotation von Sprachdaten gestreift: Die bestimmte Flexionsform eines Lemmas kann entscheidend für die Typik eines Ausdrucks sein, was zeigt, wie wichtig es sein kann, mit Token statt mit Lemmata zu arbeiten (vgl. Kapitel 6.5).

muss ein statistisches Modell zur Berechnung solcher Ausdrücke berücksichtigen. Doch ist das Phänomen überhaupt statistisch fassbar? Hierzu noch einmal Feilke (2003, 214) im Zusammenhang mit den idiomatischen Prägungen im Wetterbericht:

Tritt nun zu *Durchzug von* wiederum in arbiträrer und konventioneller Zuordnung zu einem spezifisch meteorologischen Aussageschema das Lexem *vereinzelt* hinzu, handelt es sich mit hoher Wahrscheinlichkeit um die Textsorte „Wetterbericht“. *Weitgehend trocken* tut den gleichen Dienst. Kollokationen wie *zunehmende Niederschläge* sichern die Zuordnung auch referenziell ab. Dass es sich dabei um einen fundamental linguistischen und nicht um einen Gegenstand bloßer Statistik handelt, wird daraus ersichtlich dass etwa die Formulierungen *schlechtes Wetter* und *ziemlich nass* – unter Gesichtspunkten referentieller Bedeutung eigentlich funktional äquivalente Synonyme also –, wie entsprechende Befragungen und Korpusanalysen zeigen, kaum Jemanden auf die Idee bringen können, es handle sich um die Textsorte „Wetterbericht“. Die Koordinationsleistung liegt im Ausdruck selbst begründet. (Feilke 2003, 214)

Die Frage lautet hier, was genau „Gegenstand bloßer Statistik“ ist, oder nicht ist. Analytisch betrachtet müsste es nach meinem Verständnis gerade statistisch möglich sein, zu zeigen, dass *schlechtes Wetter* zwar für sich gesehen eine überzufällig vorkommende Wortkombination ist, deren Vorkommen zusammen mit *Durchzug von* oder *vereinzelt* sehr unwahrscheinlich ist. Bei der weiteren Analyse dürfte sich zeigen, dass es Cluster gibt, die aus überzufällig häufig gemeinsam auftretenden Ausdrücken bestehen und deshalb eine spezifische Kontextualisierungsleistung erbringen, also z. B. typische Ausdrücke der Textsorte ‚Wetterbericht‘ umfassen.

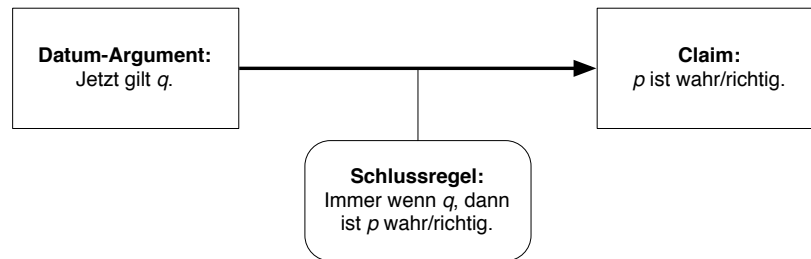
#### 4.5 Argumentationsfiguren, Topoi

Die Argumentationsanalyse versucht die rhetorische Gestaltung und Funktion von Texten zu beschreiben. Es existiert eine lange Tradition, die Argumentation als ein Element neben anderen in die Rhetorik einbettet und deren Wurzeln in die Antike reichen.<sup>13</sup>

„Am Anfang einer Argumentation steht eine strittige Aussage“ (Eggler 2006, 5).<sup>14</sup> Argumentation „als methodisches Verfahren ra-

<sup>13</sup> Vgl. hierzu Kienpointner (1992) und Kopperschmidt (2000).

<sup>14</sup> Oder um Quintilian zu zitieren: „argumento autem nisi in re controversia locus esse non potest“ (zit. nach Kopperschmidt 1989, 14; Quintilians „Institutio oratoria“ V 9,1).



**Abbildung 4.1:** Das einfache Toulminschema (vgl. Toulmin 1958 und Eggler 2006, 27).

tionaler Problembewältigung“ (Kopperschmidt 1989, 120) dient dem Zweck, Geltungsansprüche, die eine Aussage erhebt, einzuklagen oder einzulösen. Es stehen also im Raum eine Behauptung  $p$ , ein Argument  $q$  und eine Schlussregel  $SR$ , die begründet, weshalb aus dem möglichst unstrittigen Argument  $q$  die Behauptung  $p$  zwingend folgen muss:

$$p \text{ ist wahr/richtig, weil } q - \text{gemäß } SR \quad (4.1)$$

In Anschluss an Toulmin (1958) lässt sich dieses Grundschema einer Argumentation in die Positionen ‚Behauptung‘/, ‚Claim‘, ‚Datum‘ und ‚Schlussregel‘/, ‚Warrant‘ gießen und in die grafische Form in Abbildung 4.1 bringen.

Ein Beispiel:

- (9) ich behaupte mal, dass dein bsp. wegen dem scanf nicht funktioniert..du kannst einen double als float ausgeben, aber nicht einen float einlesen und als double speichern...hatte glaub ich genau dieses problem auch mal.. *Inforum* → *Bachelor* → *Bachelor 2. Jahr* → *Systemnahe Programmierung* → *cycle counter*: *Autor incubus*, 14. 2. 2007, 17:58 Uhr.<sup>15</sup>

Das ist eine Antwort in einem Forum, in dem sich Informatik-Studierende austauschen. Auch wenn der Inhalt fremd ist, zeigen sich die Positionen des Toulminschemas:

<sup>15</sup> Vgl. <http://forum.vis.ethz.ch/archive/index.php/t-9279.html> (14. 2. 2008).

*Behauptung/Claim:* Das Beispiel funktioniert wegen der falschen Verwendung von ‚scanf‘ nicht.

*Datum-Argument:* Es wurde mit ‚scanf‘ ein ‚float‘ eingelesen und als ‚double‘ gespeichert.

*Schlussregel:* Man kann mit ‚scanf‘ einen ‚double‘ als ‚float‘ einlesen, aber nicht umgekehrt.

Neben diesen Positionen ist sogar noch eine sog. ‚Backing‘-Position (Eggler 2006, 27) erfüllt: Die Erfahrung „hatte glaub ich genau dieses problem auch mal“ dient nämlich der Stützung der Schlussregel und macht sie dadurch plausibler.

Einer der wichtigsten Aspekte der Argumentationsanalyse ist die Beschreibung des musterhaften Gebrauchs von Argumenten. Die Topik, zurückgehend auf Aristoteles, Cicero und Quintilian, versucht typische ‚Fundorte‘ für wirkungsvolle Argumentationen zu erfassen. Topoi sind demnach (mehr oder weniger abstrakt definierte) Schlussregeln, die auf alltagslogischen Denkmustern oder konventionellem Erfahrungswissen beruhen, so z. B. Kausalschlüsse, Vergleiche, Einordnungen usw.

Im Kontext dieser Arbeit interessiere ich mich primär für textlinguistische Zugriffe, die Argumentation, oder besser: Argumentationsmuster als Indikatoren für die Funktionsweise von Diskursen auffassen. Damit liegt die Topik im Zentrum des Interesses, denn sie versucht gerade das Musterhafte von Argumentation zu fassen.

Im Bereich der Linguistik scheinen mir hier methodisch zwei Arbeiten richtungsweisend zu sein: Die Untersuchung von Wengeler (2003) zu Topoi und Argumentation im Migrationsdiskurs von 1960–1985 und die Analyse von Eggler (2006): „Argumentationsanalyse textlinguistisch“, die argumentative Figuren für und wider den Golfkrieg von 1991 zum Thema hat. Die beiden Arbeiten vertreten methodisch die Pole einer Skala: Wengeler (2003, 340) untersucht knapp 1400 Zeitungsartikel, Eggler (2006, 1) deren zwei. Ersterer zählt und klassifiziert in den Texten vorkommende Topoi und versucht auf deren Basis eine Mentalitätsgeschichte zu schreiben, letzterer rekonstruiert das argumentative Gerüst zweier Leitartikel, die für und wider den gleichen Gegenstand argumentieren, und versucht daraus Regeln über das Argumentieren in Ausnahmesituationen abzuleiten.

Dreh- und Angelpunkt beider Zugänge sind Topoi. Mit dem Toposbegriff könnte man „leicht in die Gefahr [geraten], sich im begriffsgeschichtlichen Gestrüpp zu verfangen oder der Versuchung [erliegen], sich durch dogmatische Begriffsnormierungen gewaltsam einen Weg durch dieses Gestrüpp zu schlagen“; doch der Streit um das angemessene Topikverständnis ist „nicht zu schlichten[,] [...] weil er mittlerweile selbst Teil der Begriffsgeschichte der Topik geworden ist“ (Kopperschmidt 1991, 53). Das Problem liegt darin, dass sich grob zwei Lesarten von ‚Topos‘ entwickelt haben:

1. Topoi als konventionalisierte Denk-, Wertungs- und Urteilmuster, die „Orientierung für die Bewältigung [der] gesellschaftlichen Existenz“ bieten (Kopperschmidt 1989, 175f.).
2. Topik als „System von Strukturmustern bzw. Formprinzipien möglicher Argumente“ (Kopperschmidt 1989, 188). Damit ist Topik „eine *Heuristik* möglicher Argumente“, die „Anweisungen für das systematische Auffinden von Argumenten“ bereit stellt (Kopperschmidt 1989, 189).

Kopperschmidt (1989) bettet nun diese beiden Lesarten des Toposbegriffs in ein umfassenderes System der Argumentationsanalyse ein, die zwei Ebenen von argumentativen Texten umfasst: Eine Mikro- und eine Makroebene. Ich möchte im Folgenden die Systematik der vorgeschlagenen Analyse kurz skizzieren:

#### 1. Mikrostrukturelle Argumentationsanalyse<sup>16</sup>

##### a) Funktionale Argumentationsanalyse

Wie wird in einer Argumentation erreicht, dass eine Behauptung p durch ein Argument q überzeugungskräftig gestützt wird?

##### b) Materiale Argumentationsanalyse

Argumentative Äußerungen finden in einem kategorialen Sprachsystem<sup>17</sup> SY statt: p, weil q aufgrund der Schlussre-

<sup>16</sup> Ich folge hier Kopperschmidt (1989, 123ff.).

<sup>17</sup> Der Terminus ‚Sprachsystem‘ kann im linguistischen Zusammenhang missverständlich sein. In der Argumentationsanalyse ist mit ‚Sprache‘ die materiale Grundlage des Argumentierens gemeint, die (neben Sprache) auch Denksysteme, Ideologien und ähnliches enthält.

gel SR in SY. So gelten z. B. im System ‚Wirtschaft‘ andere Schlussregeln als im System ‚Religion‘. Solche Sprachsysteme können unterschiedliche Argumentationsgrundsätze aufweisen oder Argumente unterschiedlich gewichten. Diese ‚materiale Definition‘ einer Argumentation wird in diesem Analyseschritt rekonstruiert.

c) Formale Argumentationsanalyse

Untersuchung der Argumente hinsichtlich ihrer Muster, nach denen sie strukturell gebildet sind. Daraus folgt eine Typologisierung von Argumenten nach abstrakten Formprinzipien (z. B. ‚Gegensatzschema‘ vs. ‚Kausalschema‘), auf die sich ihre Struktur zurückführen lässt.

2. Makrostrukturelle Argumentationsanalyse<sup>18</sup>

Analyse komplexer Argumentationen, die mehrere (eingliedrige) Argumentationsstränge enthalten und so zu mehrsträngigen, komplexen Argumentationen werden, oder die nur eine eingliedrige, jedoch komplexe Argumentation enthalten. Weiter ist eine Unterscheidung in konvergente und kontroverse Argumentationen vorgesehen (Kopperschmidt 1989, 210).

Das obige System macht nun deutlich, dass zwei Arten von Topik unterschieden werden können. Die materiale Topik, die unter der Perspektive der materialen Argumentationsanalyse untersucht wird, ist stärker an spezifische Kontexte gebunden. Wengeler (2003, 266) und Egger (2006, 35) bezeichnen diese im Anschluss an den aristotelischen Toposbegriff auch als „besondere“ oder „spezifische Topoi“.

Die formale Topik dagegen wird unter der Perspektive der formalen Argumentationsanalyse untersucht und beschreibt auf einer abstrakteren Ebene Muster von Schlussregeln. Alternativ können hier auch die Bezeichnungen „allgemeine Topoi“ (Wengeler 2003, 265) oder „strukturelle Topoi“ (Egger 2006, 35) verwendet werden.

Die Unterscheidung zwischen diesen beiden Topostypen ist allerdings im Einzelfall nicht immer einfach – und auch nicht unbedingt nötig. So liegt für Wengeler (2003, 268) mit dem Interesse an mentalitätsgeschichtlichen Charakterisierungen der Reiz im Zwischenbereich von formalen und materialen Topoi. Denn für ihn geht es darum,

<sup>18</sup> Ich folge hier Kopperschmidt (1989, 206ff.).

[d]ie Strukturen bzw. Muster von Argumentationen, die explizit oder implizit in Texten vorhanden sind, zu analysieren und für verschiedene Zeitspannen zu vergleichen. Dabei sollen aber nicht die ganz allgemeinen Formprinzipien, sondern an diese Prinzipien angelehnte, mit Inhalten aus dem thematisch bestimmten Diskurs gefüllte Muster erschlossen werden. (Wengeler 2003, 268)

Dafür erweitert Wengeler (2003, 268) den Topikbegriff und bezieht die Typologie von Kienpointner (1992, 47) mit ein, der den Anspruch erhebt „die in geschriebenem und gesprochenem Standard-Deutsch der Gegenwart anzutreffenden Argumentationsschemata annähernd vollständig zu erfassen“ (Kienpointner 1992, 47).

Abbildung 4.2 auf der nächsten Seite zeigt die Typologie der Topoi der Alltagslogik von Kienpointner (1992, 246) im Überblick. Für jedes Argumentationsschema werden verschiedene Ausprägungen aufgeführt und mit realen Textausschnitten illustriert, so dass insgesamt rund sechzig Argumentationsschemata differenziert werden. Darunter findet sich z. B. ein Typus des Vergleichsschemas, das ‚Gerechtigkeits-schema‘, das über die Kategorie der Gerechtigkeit als Schlussregel funktioniert:

Wenn X und Y hinsichtlich eines quantitativen/qualitativen Kriteriums Z gleich/ähnlich sind, sind sie im Normalfall diesbezüglich gleich/ähnlich zu bewerten bzw. zu behandeln.

X und Y sind hinsichtlich Z gleich/ähnlich.

Also: X und Y sind im Normalfall hinsichtlich Z gleich/ähnlich zu bewerten/zur behandeln. (Kienpointner 1992, 286)

Kienpointner (1992, 288ff.) präsentiert aus seinen Korpusdaten mehrere Beispiele für die Anwendung dieses Topos, beispielsweise im Rahmen einer Diskussion um die Militärdienstpflicht für Frauen,<sup>19</sup> aber auch Varianten, wie die in der Werbung gebräuchliche, in der die Eigenschaften eines positiv konnotierten Vergleichsgegenstands

<sup>19</sup> Hier in zweifacher Anwendung in der Diskussion zwischen einer Managerin I. Moser und der Feministin A. Schwarzer. Das Argument von Moser:

**Datum:** Männer müssen in die Armee, Frauen nicht.

**Schlussregel:** Frauen und Männer sind gleich.

**Behauptung:** Auch Frauen müssen Militärdienst leisten.

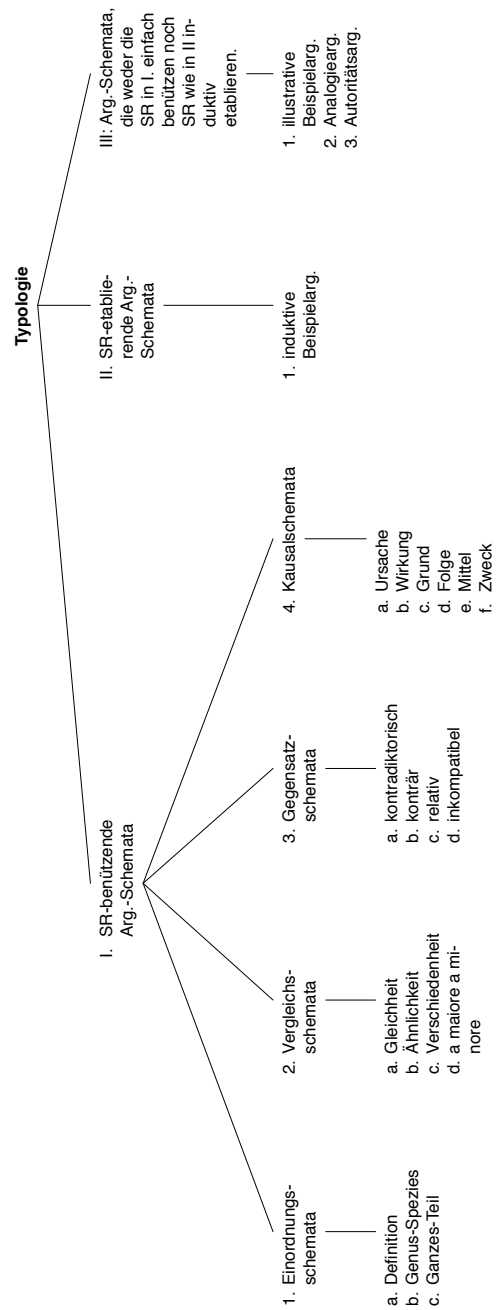
Die Gegenargumentation von Schwarzer:

**Datum:** Frauen übernehmen mehr/ebenfalls Pflichten, die Männer nicht übernehmen.

**Schlussregel:** Frauen und Männer sind gleich.

**Behauptung:** Wenn von Gerechtigkeit die Rede sein soll, dann müssten zuerst die Männer nachholen und Pflichten übernehmen, die Frauen einseitig übernehmen müssen.

(Kienpointner 1992, 295f.)



**Abbildung 4.2:** Die Typologie der Topoi der Alltagslogik (Kienpointner 1992, 246).



als Datum-Argument für die Güte des beworbenen Produkts dienen sollen: „Timotei Shampoo wäscht Ihr Haar so mild wie die Natur“ (Kienpointner 1992, 304).

Diese ausdifferenzierten Schemata sind alle eher abstrakter Natur und kaum themengebunden; es handelt sich also um formale bzw. allgemeine/strukturelle Topoi. Die beiden Arbeiten von Wengeler (2003) und Eggler (2006) zeigen jedoch, dass eine Beschränkung auf Topoi dieses Typus nicht sinnvoll ist, wenn das Untersuchungsinteresse einem Diskurs bzw. der konkreten Argumentation zu einem Thema gilt.<sup>20</sup> So macht Wengeler (2003, 276f.) deutlich, dass bei der Bewusstseins- und Mentalitätsgeschichtsschreibung sowohl formale Topoi, jedoch auch ihre „inhaltliche, materiale Füllung“, also die materialen Topoi, die Begründungssprachen, von Interesse seien. Auch Eggler (2006, 35f.) erfasst sowohl strukturelle Topoi<sup>21</sup> als auch spezifische Topoi.<sup>22</sup>

Der argumentationstheoretische Zugang zu Texten, wie ihn Kopperschmidt (1989) skizziert, wie er für Kienpointner (1992) als Basis zur Erarbeitung einer Typologie der Topik im Gegenwartsdeutsch dient und wie ihn auch die linguistischen Arbeiten von Eggler (2006) und Wengeler (2003) wählen, ist durch die folgenden beiden Momente geprägt:

1. Die Analysen erfolgen mehr oder weniger deduktiv: Es wird von bestehenden Typologien der Topik ausgegangen, die aufgrund von Korpusdaten (im Umfang zwischen mehreren hundert und ‚nur‘ zwei Texten) auf ihre Verwendung hin überprüft werden.
2. Die Analysen fokussieren die semantische Tiefenstruktur: Das Analyseverfahren ist hermeneutisch. Die Texte werden semantisch entschlüsselt, also interpretiert. Auf diese Weise werden

<sup>20</sup> Dies ist auch Kienpointner (1992, 235) klar, der seine Typologie als Grundlage für andere Untersuchung sieht, die „jedoch thematisch-kontextuell weit mehr ins Detail gehen [können], wenn etwa spezielle Argumentationsformen vor Gericht untersucht werden oder die Argumentation einer bestimmten politischen, religiösen oder generations- bzw. geschlechtsspezifischen Subgruppe näher analysiert werden soll, oder wenn die Argumentation einer Einzelperson beschrieben werden soll“.

<sup>21</sup> Wie beispielsweise das „Argumentum ex auctoritate“, das „Argumentum per analogiam“ oder das „Argumentum a fortiori“ (Eggler 2006, 42ff./56ff./60ff.).

<sup>22</sup> Ein Beispiel aus der Diskussion um den Golfkrieg: „Wenn ein Diktator vom Schlage Hitler eine ganze Weltgegend bedroht, muss man ihn stoppen“ (Eggler 2006, 36).

Argumentationsmuster auf der semantischen Tiefenstruktur rekonstruiert.

Solche Analysen sind für einen induktiven, korpuslinguistischen Zugang, der möglichst umfangreiche Textmengen bewältigen soll, problematisch. Denn es stellt sich die Frage nach der Operationalisierung von Topiktypologien, die nicht auf der semantischen Tiefenstruktur, sondern auf der Textoberfläche liegen müssten. Denn nur so ist eine (halb-)automatische Verarbeitung von Textdaten und damit ein induktiver Zugang möglich.

Doch würde eine Operationalisierung bestehender Topiktypologien mit dem Ziel einer induktiven Analyse diese ad absurdum führen. Denn mit einer induktiven Analyse läge die Chance ja gerade darin, bestehende Typologien zu hinterfragen und durch eine induktiv aus den Daten abgeleitete Typologie zu ergänzen.

Trotzdem scheint der Topos eine fruchtbare Kategorie zu sein, um die Sprechweisen in Diskursen fassen zu können. Allerdings ist es für eine induktive Analyse nicht sinnvoll, bestimmte (formale oder materiale) Topoi zu operationalisieren, sondern das Konzept des Topos ganz allgemein muss operationalisiert werden. Oder anders gewendet: Bei der induktiven Analyse von musterhaften Strukturen im Korpus muss klar werden, welche Muster nun topisch sind und welche nicht.

Und hier komme ich wieder an den Anfang dieses Kapitels zurück: Textmaterial, um es so offen wie möglich zu formulieren, ist dann topisch, wenn es musterhaft ist und für einen argumentativen Zweck verwendet wird, wenn es also möglich ist, ein Toulminschema zu zeichnen, in dem das topische Muster als Schlussregel dient. Diese letzte Bedingung, dass das topische Muster die Position der Schlussregel einnehmen muss, dürfte zu strikt sein. Es ist nicht zu erwarten, dass Schlussregeln oft verbalisiert werden.<sup>23</sup> Daher muss Topik etwas weiter gefasst werden, als darunter nur musterhafte Schlussregeln zu verstehen. Ich werde auch Muster als topisch bezeichnen, die an anderen Positionen eines Toulminschemas Verwendung finden, also z. B. (Teil-)Verbalisierungen des Datum-Arguments oder des Claims sind. So würde ich den Sprachgebrauchsmustern X WARNT VOR SALAMITAKTIK, bzw. dem dazu alternativen Ausdruck ERST X, DANN Y

<sup>23</sup> Vgl. Egger (2006, 79).

(vgl. unten Punkt 1), und IST/HAT NUN EINMAL (vgl. unten Punkt 2) topische Funktion zuschreiben:

1. X WARNT VOR SALAMITAKTIK

ERST X, DANN Y Diese beiden Sprachgebrauchsmuster können in topischer Verwendung auf die folgende Schlussregel zurückgeführt werden: „Wenn erstmal ein Schritt in eine bestimmte Richtung gemacht wird, geht damit die ganze Entwicklung in dieser Richtung weiter“ ‚Slippery-Slope-Argument‘.

Im folgenden Beispiel werden beide Formulierungen im selben Text verwendet:

(10) **„Schaar warnt vor Salamitaktik**

Der deutsche Bundesdatenschützer Peter Schaar hat in Sachen Online-Durchsuchung vor einer Salamitaktik der Behörden gewarnt. [...]

**Erst die Polizei, dann das Finanzamt**

Ähnlich sei es auch beim Verfahren zum Kontenabruf gewesen, das ursprünglich dazu dienen sollte, terroristische Geldströme aufzudecken: „Heute haben jedes Finanzamt und die Sozialbehörden Zugriff auf diese Daten. Ohne die Begründung Terrorbekämpfung wäre diese Abfragemöglichkeit aber überhaupt nicht durchsetzbar gewesen.“ *Futurezone.ORF.at*, 25. Oktober 2007.<sup>24</sup>

2. IST/HAT NUN EINMAL

Dieses Sprachgebrauchsmuster kann in topischer Verwendung auf eine Schlussregel zurück gehen, die vom Faktischen auf das Normative schließt (‚naturalistischer Fehlschluss‘).

Beispiel:

(11) **„Blickpunkt: Wie steht die PDS dazu, Frau Maier? Darf sich der Staat aus seiner Verantwortung für die Arbeitsvermittlung verabschieden?**

<sup>24</sup> Vgl. <http://futurezone.orf.at/it/stories/231122/> (1. 11. 2007).

Pia Maier: Nein, das darf er nicht. Der Weg, die Privatvermittlung noch weiter zu stärken, ist der falsche Weg. Denn er funktioniert einfach nicht. Arbeitsvermittlung **ist nun einmal** kein marktfähiges Geschäft.“ *Blickpunkt* 2/2002.<sup>25</sup>

Mit einer solchen Definition wird aber auch klar, wo die Grenzen eines induktiven Zugangs liegen. Die Kategorie ‚Argumentation‘, die zu einem Toulminschema verdichtet werden kann, muss nicht aufgegeben werden. Es werden sich auch nicht völlig neue formale Topoi finden lassen. Hingegen wird interessant sein, zu sehen, wie die Topoi materialisiert werden und in welchen Kontexten sie auftauchen.

Die Frage nach solchen Indikatoren für Argumentationsfiguren auf der Textoberfläche wird im Kontext der Argumentationsanalyse immer wieder thematisiert. So z. B. von Kopperschmidt (1989, 99/64ff.), der von „argumentativen Operatoren“ spricht. Auch Kienpointner (1992, 237f.) macht auf solche Indikatoren wie Metasprache (*Lassen Sie mich einen Analogieschluss gebrauchen...*) und lexikalisches Material (*erst, erst recht, schon*) aufmerksam. Es ist jedoch klar, dass diese sprachlichen Ausdrücke alleine nicht zuverlässig als Indikatoren funktionieren, um zweifelsfrei Argumentationen auszumachen. So wären die sprachlichen Muster *erst X, dann Y* oder *ist/hat nun einmal* auf der vorherigen Seite auch in nicht-topischen Kontexten möglich.

Um zuverlässige Indikatoren zu finden, bietet sich der umgekehrte Weg an: Mit einem induktiven Zugang eröffnet sich das Potenzial, für bestimmte Teilkorpora typische Indikatoren auf der Textoberfläche zu finden, aus denen Topoi abgeleitet werden können. Damit müsste die Möglichkeit bestehen, doch noch eine Operationalisierung von spezifischen Topoi zu leisten, die auf die Textoberfläche zugreift. Darin liegt wahrscheinlich der größte Vorteil einer induktiven Analyse.

## 4.6 Metaphern

Metaphern sind nicht einfach ein eher seltenes Phänomen literarischer oder rhetorischer Rede. Ganz im Gegenteil strukturieren sie als

<sup>25</sup> Vgl. <http://www.bundestag.de/bp/2002/bp0202/0202092a.html> (1.11.2007).

‚konzeptionelle Metaphern‘ den Sprachgebrauch stark, wie Lakoff/Johnson (1980)<sup>26</sup> gezeigt haben. Am Beispiel des Sprechens über Krieg illustriert Lakoff (1991, 221) dies wie folgt:

Das metaphorische Verständnis einer Situation funktioniert in zwei Abschnitten. Erstens gibt es eine umfassende, relativ konstante Anzahl von Metaphern, die unser Denken strukturieren. Beispielsweise könnte man die Entscheidung, einen Krieg zu beginnen, als eine Form von Kosten-Nutzen-Analyse betrachten, die einen Krieg rechtfertigt, wenn die Kosten dafür geringer sind als die Kosten, keinen Krieg zu beginnen. Zweitens gibt es eine Reihe metaphorischer Definitionen, die es einem erlauben, eine solche Metapher auf eine bestimmte Situation anzuwenden. In diesem Fall muß es für „Kosten“ eine Definition geben, einschließlich einer Möglichkeit, einen Vergleich der relativen „Kosten“ anzustellen. Die Anwendung einer Metapher mit einer Reihe von Definitionen wird dann bedenklich, wenn sie die Tatsachen auf gefährliche Art und Weise verschleiert. (Lakoff 1991, 221)

Am Beispiel der politischen Diskussionen vor dem Golfkrieg von 1991 zeigt Lakoff (1991), wie der Diskurs in den USA durch konzeptionelle Metaphern wie „Krieg als Politik“ (‚Clausewitzsche Metapher‘), „Politik als Geschäft“ oder mit dem „System des kausalen Handelns“ vorstrukturiert wurde. Solche Metaphern erlauben Folgemetaphern, die Krieg als rationale Rechnung von Kosten und Gewinn, Risiko, Profit und Zweckmäßigkeit sehen lassen. Dabei werden andere mögliche Metaphern in den Hintergrund gedrängt, z. B. „Krieg als Gewaltverbrechen: Mord, Überfall, Entführung, Brandstiftung, Vergewaltigung und Diebstahl“ Lakoff (1991, 229).

Das Resultat solcher Strukturierungen durch Metaphern zeigt sich auch auf der lexikalischen Ebene. Dies beispielsweise in Einzellexemen wie *Kosten*, *Nutzen*, *(Spiel-)Einsatz*, aber auch in Wendungen wie *seine Karten offen auf den Tisch legen*, die alle als Effekte der Metapher ‚Krieg als rationales, kausales Handeln‘ oder ‚Risiko = Glücksspiel‘ gesehen werden können. Detaillierter zeigen dies Lakoff/Johnson (1980) für eine Reihe von konzeptionellen Metaphern in „Metaphors we live by“ auf, wo klar wird, wie umfassend die Autoren die strukturierende Kraft von Metaphern annehmen, wenn von Metaphern wie ‚Zeit ist Geld‘, ‚Liebe ist Verrücktheit‘, ‚Theorien sind Gebäude‘, aber auch ‚gut ist oben; schlecht ist unten‘ oder ‚die Inflation ist eine Person‘ die Rede ist.

<sup>26</sup> Dt.: Lakoff/Johnson (1998).

Kontrovers wird in diesem Zusammenhang diskutiert, ob das Primat des Kognitiven oder des Sprachlichen gilt:<sup>27</sup> Sind es die kognitiven Konzepte (konzeptionelle Metaphern), die die sprachlichen Ausdrucksmöglichkeiten erzeugen, wie Lakoff/Johnson (1980) dies behaupten, oder ist eher der umgekehrte Weg der Fall?

Die Konzeptualisierung, von der Lakoff/Johnson ausgehen, findet sicherlich im kognitiven Bereich statt. Doch sie ist nicht voraussetzungslos. Die Konzeptualisierung [...] setzt nicht unbedingt Erfahrungen im Sinne persönlicher Bekanntheit voraus, es genügen *Beschreibungen* [...]. Solche Beschreibungen jedoch sind sprachlich kommunizierte Erfahrungen. (Rolf 2005, 241)

Im Rückgriff auf Feilke (1993) (vgl. Kapitel 4.4) würde man den Zusammenhang zwischen Kognition und Sprache wohl als wechselseitig charakterisieren. Metaphern und typische Ausdrücke überlappen sich. Metaphern zeigen sich auf der Textoberfläche teilweise als geprägte Ausdrücke, sind also Indikatoren für die Konventionalisierung von typischem sozialem Handeln. Anderes, z. B. spontane, singuläre Vergleiche, fallen hingegen nicht unter derart geprägte Ausdrücke, können aber gleichwohl als Metaphern verstanden werden. Für Metaphern des ersten Typs, die sich also in geprägten Ausdrücken zeigen, gilt sicher die enge Wechselwirkung zwischen dem durch soziales Handeln geprägtem Sprachgebrauch und dem sprachlich geprägtem sozialem Handeln.

Die neue Forschung zu Metaphern versucht nun vermehrt Korpora beizuziehen, um empirische Evidenz zu gewinnen.<sup>28</sup> Dabei steht die Entwicklung von Methoden im Vordergrund, um möglichst maschinell Metaphern in Korpora aufzuspüren.<sup>29</sup> Stefanowitsch (2006a, 2ff.) unterscheidet dabei (neben der manuellen Suche und der Suche in Korpora, die mit Metapherninformationen annotiert sind) eine Reihe

<sup>27</sup> Vgl. zu dieser Diskussion Rolf (2005, 240f.).

<sup>28</sup> Für einen Überblick über die aktuelle Forschung vgl. Stefanowitsch/Gries (2006).

<sup>29</sup> Daneben sind natürlich weitere Perspektiven auf Metaphern in Korpora möglich. Müller (2006) benutzt Korpora z. B. dafür, sog. kreative Metaphern, deren Existenz häufig nur für literarische Texte behauptet wird, auch in politischen Texten nachzuweisen oder die Innovation einer Metapher zu prüfen. Daneben gibt es unzählige Arbeiten, die im Rahmen umfassender Diskursanalysen mit dem Metaphernkonzept arbeiten. So beispielsweise Burkhardt (2003) im Rahmen seiner Analyse der Sprache des Parlaments, Geideck/Liebert (2003), die „Sinnformeln“ teilweise in Metaphern realisiert sehen, Musolf (2005) zur argumentativen Funktion von Krankheitsmetaphern oder Spitzmüller (2005), der den Metasprachdiskurs zu Anglizismen auch durch die Beobachtung von Metaphern charakterisiert.

von Ansätzen, bei denen meist das Vokabular des Quell- und/oder Zielbereichs einer Metapher Ausgangspunkt der Suchmethoden ist. Bei der Metapher *die EU ist das Haus Europas* wäre EU Ziel- und Haus Quellvokabular. Folgende Suchstrategien können angewendet werden:

1. Suche nach Vokabular eines Quellbereichs (im Kontext der EU z. B. nach *Haus*, *Fundament* etc.)
  - a) mittels apriori festgelegter Suchbegriffe,
  - b) mittels ganzer Wortlisten<sup>30</sup> oder
  - c) auf Basis einer vorherigen Schlüsselwortanalyse.<sup>31</sup>

Anschließend werden die Zielbereiche, und damit die Metaphern, identifiziert. Im Kontext der EU könnte *die Verfassung* (Zielbereich) als *Fundament* beschrieben werden; die Metapher lautete dann: DIE EU-VERFASSUNG IST DAS FUNDAMENT DER EU.

2. Suche nach Vokabular des Zielbereichs (im Kontext der EU z. B. *EU*, *Europäische Union*, *Brüssel* etc.)
  - a) mittels apriori festgelegter Wortlisten oder
  - b) auf Basis einer vorherigen Schlüsselwortanalyse.

Anschließend werden die Metaphern aufgrund des typischen Quellbereichsvokabulars in der Umgebung der gefundenen Wörter des Zielbereichs identifiziert. Im Kontext der EU könnte *disziplinierendes Korsett*<sup>32</sup> oder *Fahrrad*<sup>33</sup> Vokabular aus dem

<sup>30</sup> Koller (2006, 244ff.) untersucht beispielsweise Zeitungstexte zu den Themen Marketing und Vertrieb, sowie Fusionen und Übernahmen von Unternehmen. Um die dort vorherrschenden Metaphern zu finden, arbeitet sie mit je 105 Lemmata aus je drei Quellbereichen (Krieg; Sport und Spiel/evolutionäres Kämpfen; Paaren und Ernähren). In der Korpusanalyse sucht sie nach diesen Lemmata und kann so quantitative Aussagen zur Verwendung von Metaphern im Korpus machen.

<sup>31</sup> Zu diesem Verfahren vgl. die Ausführungen weiter unten und Partington (2006).

<sup>32</sup> Vgl. den Artikel „Jean-Claude Juncker: ‚Die EU ist ein disziplinierendes Korsett‘. Der luxemburgische Premierminister über die EU-Erweiterung“ in der Frankfurter Rundschau vom 16. Februar 2004: [http://www.gouvernement.lu/salle\\_presse/interviews/2004/02/20040216juncker\\_fr/index.html](http://www.gouvernement.lu/salle_presse/interviews/2004/02/20040216juncker_fr/index.html) (3. Juli 2008).

<sup>33</sup> Vgl. den Artikel „Klapprad ohne Gänge“ von Eric Bonse auf Handelsblatt.com vom 23. Juni 2008: „Europa ist wie ein Fahrrad, pflegte der frühere EU-Kommissionschef Jacques Delors zu sagen: Wenn es zum Stehen

Quellbereich sein. Die Metaphern lauteten dann: DIE EU IST WIE EIN KORSETT bzw. DIE EU IST WIE EIN FAHRRAD.

3. Kombinierte Suche nach Vokabular von Quell- und Zielbereich.
4. Suche nach Metaphern basierend auf Markern, wie *metaphorisch gesprochen, sozusagen, Bild, Ähnlichkeit mit...* etc.
5. Extraktion von Metaphern aus einem semantisch annotierten Korpus.

Doch diese Suchstrategien sind nur der erste Schritt. Denn die Identifikation der Metaphern, also die Entscheidung, ob und welche Art von Metapher in einer Textstelle verwendet wird, basiert „on more-or-less explicit commonsensical intuitions of the part of the researcher“ (Stefanowitsch 2006a, 10). Nur klare Kriterien, verbunden mit Reliabilitätstests, könnten dieses Problem entschärfen.

Stefanowitsch (2006b) zeigt mit der vorgeschlagenen „Metaphorical Pattern Analysis“ (MPA) wie die konzeptionellen Metaphern bestimmter Ausdrücke des Zielbereichs durch eine Korpusanalyse erarbeitet werden können. Allerdings schränkt Stefanowitsch (2006b, 66) den Metaphernbegriff leicht ein, indem er als metaphorisches Muster („metaphorical pattern“) definiert: „A metaphorical pattern is a multi-word expression from a given source domain (SD) into which one or more specific lexical item from a given target domain (TD) have been inserted.“ (Stefanowitsch 2006b, 66) Damit wird eine MPA möglich:

Crucially, metaphorical patterns provide a basis for target-domain oriented studies on the basis of corpus data: we can retrieve a large number of instances of a target domain item (such as *claim, criticism, argument*, etc.) from a corpus and exhaustively identify the metaphorical patterns that it occurs with. Obviously, this kind of procedure, which I will refer to as *metaphorical pattern analysis* (MPA) will capture only a subset of metaphorical expressions – those manifesting themselves as metaphorical patterns for specific lexical items [...]. (Stefanowitsch 2006b, 66)

Die MPA geht also davon aus, dass ein solches Metaphernmuster sich in Wörtern des Quell- und des Zielbereichs lexikalisiert. So wird im Ausdruck *sie kocht vor Wut* mit *Wut* der Zielbereich ‚Emotionen‘, mit *kochen* der Quellbereich ‚erhitzte Flüssigkeit‘ lexikalisiert und

---

kommt, kippt es um.“: <http://www.handelsblatt.com/politik/handelsblatt-kommentar/klapprad-ohne-gaenge;1446655> (3. Juli 2008).



dadurch kommt der Vergleich ‚wütend sein = heiße Flüssigkeit in einem Behälter sein‘ zu Stande.

In seiner Analyse zu Metaphern für Emotionen sucht Stefanowitsch (2006b, 70ff.) nun in einem Korpus nach Wörtern des Zielbereichs, nämlich nach Ausdrücken wie *anger*, *disgust*, *fear*, *happiness* und *sadness* und analysiert und kategorisiert diese. Das erlaubt, die Resultate zu quantifizieren und z. B. die Verwendung dieser Metaphern im Untersuchungskorpus mit dem Gebrauch in anderen Korpora zu vergleichen. Weiter ist es möglich, die Signifikanz der Korrelation zwischen den Metaphern und den lexikalisierten Emotionen zu berechnen, um abschätzen zu können, welche Metaphern für die jeweiligen Emotionen typisch sind.

Stefanowitsch (2006b, 70ff.) vergleicht das Verfahren der MPA mit der klassischen „introspektiven“, qualitativ-hermeneutischen Methode, die in der kognitiven Linguistik üblicherweise verwendet wird.<sup>34</sup> Für diesen Vergleich dient eine Arbeit von Kövecses (1998), in der der Autor die in der Forschungsliteratur bestehende Sammlung von Metaphern für Emotionen auszuweiten versucht. Im Vergleich zu dieser Sammlung findet Stefanowitsch (2006b, 90) nicht nur die große Mehrheit der in der Arbeit von Kövecses (1998) versammelten Metaphern, sondern darüber hinaus ebenso viele Metaphern, die in traditionellen Analysen nicht gefunden worden sind.

Einen anderen Weg wählt Partington (2006), der durch den Vergleich von mehreren Korpora Schlüsselwortlisten erstellt, die für die verglichenen Korpora die je typischen Wortformen enthalten. Diese sind der Ausgangspunkt für weitere Analysen, bei denen auch signifikante Kollokationen zu den Schlüsselwörtern berechnet werden. In Kombination mit qualitativen Methoden „including intuition, introspection and immersion in a text“, skizziert Partington (2006, 299) eine Methodik der „Corpus-Assisted Discourse Studies“, bei der jedoch die Korpusanalyse „strictly functional to the overall task in hand“ sein soll.

Das Konzept der Metapher ist ein Konstrukt auf der semantischen Tiefenstruktur. Nur teilweise können Metaphern durch Indikatoren

<sup>34</sup> Die Bezeichnung ‚introspektiv‘ für diese Methoden ist m. E. unglücklich, da diese Arbeiten durchaus empirisch, also mittels Analyse von Textmaterial, vorgehen, dies jedoch eher unsystematisch, qualitativ und in einem hermeneutischen Prozess tun, bei dem die Analysen nicht nur strikt von der Textoberfläche ausgehen.

auf der sprachlichen Oberfläche gefunden werden. Aber, das zeigen die Untersuchungen von Stefanowitsch (2006b) (vgl. oben), es ist letztlich doch ein beachtlicher Teil dessen auf der Textoberfläche lexikalisch greifbar, was Metaphern-Forscherinnen und Forscher als Metaphern bezeichnen würden.

Für mein Untersuchungsinteresse müssen folgende Fragen im Zusammenhang mit den dargestellten Metaphern-Konzepten gestellt werden:

1. Hilft das Konzept der Metapher dabei, Sprechweisen in Diskursen besser beschreiben zu können?
2. Wenn ja: Können Metaphern so operationalisiert werden, dass sie induktiv-korpuslinguistisch und maschinell aus den Daten extrahiert werden können?

Die zweite Frage kann mit einem vorsichtigen Ja beantwortet werden, wie die Ausführungen oben gezeigt haben. Zwar sind viele Untersuchungen noch primär an bestimmten Metaphern in bestimmten Bereichen interessiert, doch mit den Methoden von Partington (2006) und Stefanowitsch (2006b) sind viel versprechende Ansätze da, um Metaphern maschinell aus Korpora zu extrahieren.

Die erste Frage kann ich an dieser Stelle noch nicht beantworten. Die weiteren Analysen werden zeigen, ob es sinnvoll ist, anhand der Verwendung von Metaphern bestimmte Funktionen oder Wirkungen im Diskurs zu erklären.

Doch ähnlich wie bei meinen Überlegungen zu den Argumentationsfiguren und Topoi (vgl. Kapitel 4.5) zeigt sich auch hier, dass bei einem streng induktiven Verfahren, das zunächst bloß versucht, musterhafte Strukturen in Textdaten aufzudecken, ein Konzept wie das der Metapher seine Nützlichkeit erst beweisen muss. Denn erstens ist unklar, ob die gefundenen musterhaften Strukturen überhaupt in ein solches Konzept eingepasst werden können und zweitens, eingepasst werden sollen.

#### 4.7 Sprachgebrauchsgeschichte als Kulturanalyse

Für die oben dargestellten Analysekat­egorien gilt, dass sie sowohl für synchrone als auch für diachrone Perspektiven auf Sprache verwendet werden können. Die Wahl der Perspektive ist der Wahl der Analysekat­egorie nachgelagert.

Sprachgeschichte setzt den Fokus auf die diachrone Perspektive. Die Frage nach der Veränderung von Sprache in der Zeit ist Ausgangspunkt. Die Wahl der Analysekat­egorie folgt anschließend. Für mein Vorhaben einer korpuslinguistischen Diskursanalyse ist diese diachrone Perspektive ebenfalls fruchtbar. Vor allem, wenn die neueren Entwicklungen berücksichtigt werden, mit denen sich Sprachgeschichte zu einer Sprachgebrauchsgeschichte wandelte und diese somit als Methode der Kulturanalyse an ähnlich gelagerte Forschungsfragen aus anderen Disziplinen angeknüpft werden kann.

Diesen Paradigmenwechsel zu einer Sprachgebrauchsgeschichte zeigt Linke (2003c) auf und entwirft damit ein Programm der Sprachgebrauchsgeschichte als Kulturanalyse, das sie auch mehrfach mit Einzelstudien exemplifizierte.<sup>35</sup> Kern des Programms ist im Anschluss an Cassirer die These, Sprache – genauer: Sprechen und Sprachgebrauch – als „Medium symbolischer Gestaltung“ (Linke 2003c, 44) anzusehen.<sup>36</sup>

Wenn wir Sprache als *ein*, vielleicht als *das* zentrale Symbolisierungsmedium verstehen, das es Menschen ermöglicht, in Symbolisierungsakten ihre Lebenswelt und ihr Verhalten zu dieser Welt zu gestalten und in der objektivierenden Veräußerlichung der symbolischen Formen fassbar (und damit auch kollektiv verhandelbar) zu machen, dann ist die Analyse dieses Mediums und seiner jeweiligen historischen Konkretisationen immer auch Kulturanalyse. Dass es hier im Übrigen um beides geht, um Sprachanalyse und Sprachgebrauchsanalyse, ist wichtig. Die symbolisierende Kraft von Sprache ist sowohl auf der sprachsystematischen Ebene zu verorten als auch in den Formen und Mustern des Sprachgebrauchs [...]. (Linke 2003c, 44f.)

Damit ist ein Perspektivenwechsel in der Sprachgeschichte verbunden: Statt nach den Interdependenzen zwischen einer sprachäußerlichen

<sup>35</sup> Vgl. Linke (1996, 2001, 2002, 2003b, 2004, 2006).

<sup>36</sup> Was den Paradigmenwechsel von einer Sprachgeschichte zu einer Sprachgebrauchsgeschichte betrifft, macht Linke (2003c, 30ff.) auf die bereits zahlreichen Vorarbeiten aufmerksam: So plädiert beispielsweise Cherubim (1980, 14f.) für eine historische Sprachpragmatik und schlägt dort u. a. vor, das Augenmerk besonders auf Phraseologismen und Tropen, also die musterhafte Textoberfläche, zu werfen.

Kultur und sprachlichen Formen zu suchen, wird der Sprache eine Kultur gestaltende Kraft zugeschrieben. Damit rückt die sprachliche Form als Ausgangspunkt kulturanalytischer Fragestellungen in den Fokus (Linke 2003c, 45).<sup>37</sup>

[...] der Akzent sprachhistorischer Analysen [wird] nicht (mehr) auf die Einbettung sprachlicher Phänomene in umfassendere kulturelle Zusammenhänge gelegt – in der Absicht, auf diese Weise zu einem vertieften Verständnis von Sprache und Sprachveränderung zu kommen, sondern die Analyse setzt nun an bei den sprachlichen Formen, um über die Analyse dieser Formen zu einem vertieften Verständnis der kulturellen Zusammenhänge, der kultur- und epochenspezifischen Modellierung von Erkenntnis, Erfahrung und Emotion zu kommen. (Linke 2003c, 45f.)

Mit diesem Perspektivenwechsel geht es bei der Analyse darum, ausgehend von sprachlichen Formen, „mögliche Kontexte“ (Linke 2003c, 46) zu rekonstruieren. Hier kommen Schlussverfahren wie die Griceschen Implikaturen oder Argumentationsfiguren zum Zug – oder ergänzend und allgemeiner gesagt: die ganze Fülle<sup>38</sup> von Kategorien der semantischen Tiefenstruktur, wie sie oben bereits dargestellt wurden. Natürlich rekonstruiert eine solche Analyse nur mögliche Kontexte; „je größer jedoch das Spektrum der historischen Texte ist, die wir in solche Rekonstruktionsversuche einbeziehen, desto eher lassen sich plausible von weniger plausiblen Kontexten unterscheiden“ (Linke 2003c, 46).

Doch so komplex die Konzepte auch sind, die über eine solche Kulturanalyse rekonstruiert werden, Ausgangspunkt sind sprachliche Formen und Muster. So zeigt Linke (2001) beispielsweise an Todesanzeigen auf, wie die sprachlichen Formeln, und damit die Textsorte, sich in der zweiten Hälfte des 20. Jahrhunderts grundlegend ändert. Diese Beobachtungen an der Textoberfläche deutet sie dann als Veränderungen in der Konstellation der Konzepte ‚Trauer‘, ‚Öffentlichkeit‘ und ‚Intimität‘.

In der Untersuchung zu „Spaß haben“ (Linke 2003b) ist es der Gebrauch dieser Kollokation, die im Zentrum des Interesses steht.<sup>39</sup>

<sup>37</sup> Linke (2003c, 45) legt Wert auf die Abgrenzung gegenüber einem radikalkonstruktivistisch verstandenen Konstruieren von Wirklichkeit und verwendet daher den Terminus ‚gestalten‘.

<sup>38</sup> Also auch Metaphern, Topoi, idiomatische Prägungen, diskurssemantische Grundfiguren etc.

<sup>39</sup> Erwähnenswert ist an dieser Stelle auch die Arbeit von Stocker (2005), die auch teilweise mit der Analyse von typischen Kollokationen arbeitet, um die diskursive Konstruktion des ‚Mädchens‘ im 19. Jahrhundert aufzuzeigen.

Denn „[s]eit den 80er Jahren des vergangenen Jahrhunderts ist Spaß [...] ganz allgemein zu einem gerade auch in der medialen Öffentlichkeit besonders häufig thematisierten Lebensgefühl avanciert“, wie Linke (2003b, 65) mit einer quantitativen Analyse von Zeitungsartikeln belegen kann. Und die Kollokatoren von *Spaß* verändern sich: „[...] es geht nicht mehr um *Spaß haben am Tanzen* oder allenfalls um *Spaß haben mit dem neuen Mofa*, sondern um *Spaß haben* in absoluter Verwendung: *ich habe Spaß*, *wir haben Spaß*, *sie haben Spaß*“ (Linke 2003b, 73). Verknüpft ist dieser Befund mit der Beobachtung, dass mit den 90er-Jahren die Kollokation *Spaß haben* vermehrt im kollektiven Bezug verwendet wird (*wir hatten Spaß*), und dass die Semantik sich immer weiter von einer älteren Verwendung entfernt, in der Spaß mit Ausgelassenheit und Heiterkeit verbunden wird, hin zu einem leistungsbewusst und erfolgsorientiert verstandenen *ich will Spaß haben*. Aus kulturanalytischer Perspektive deutet Linke (2003b, 78) dies wie folgt:

*Spaßhaben* in absoluter Verwendung steht nicht (mehr) im absoluten Widerspruch zu Ernsthaftigkeit und Zielstrebigkeit, sondern erscheint zumindest in Ansätzen als Ausdruck für ein Befindlichkeits-Konzept, das hedonistische Selbstbezogenheit als persönliche Leistung und als Ausweis gelungener Individuation konturiert. (Linke 2003b, 78)

Sprachgeschichte spielt sich nicht nur in Zeiträumen von Jahrzehnten oder gar Jahrhunderten ab, sondern auch innerhalb von Jahren. Dies ist die erste Grundannahme, die ich für die vorliegende Untersuchung treffen muss, wenn das Konzept der Sprachgebrauchsgeschichte dafür fruchtbar gemacht werden soll. Die zweite Grundannahme, dass die Beobachtung von Sprachgebrauch auf der Textoberfläche Ausgangspunkt für eine Analyse ist, deckt sich mit den Überzeugungen der Sprachgebrauchsgeschichte.

Ungeklärt ist an dieser Stelle das Verhältnis von Diskurs- und Kulturanalyse. Jedoch zeigt sich an den referierten Fragestellungen einer linguistischen Kulturanalyse, dass diese sich stark mit den Fragestellungen der vorliegenden Arbeit decken, die nach den Sprechweisen in Diskursen fragen. Denn die Analyse der Sprache als Symbolisierungsmedium fragt ja gerade danach, wie Symbolisierungen funktionieren, also durch welche sprachlichen Mittel auf welche Art und Weise dies geschieht.

Der Fokus auf die Analyse der (musterhaften) sprachlichen Form bietet enorme methodische Vorteile, um quantitativ, induktiv, korpus-

linguistisch zu arbeiten. Die musterhafte Verwendung von Sprachgebrauch lässt sich einfacher operationalisieren als komplexe Konzepte wie Metapher, Argumentationsfigur oder Topos. Mit quantitativen, induktiven Verfahren lassen sich Veränderungen im Sprachgebrauch nachweisen, die durch eine qualitative Analyse nicht unbedingt entdeckt würden, wenn sie nur niedrigfrequent (aber trotzdem überzufällig) sind und sich deshalb erst in großen Datenmengen zeigen.

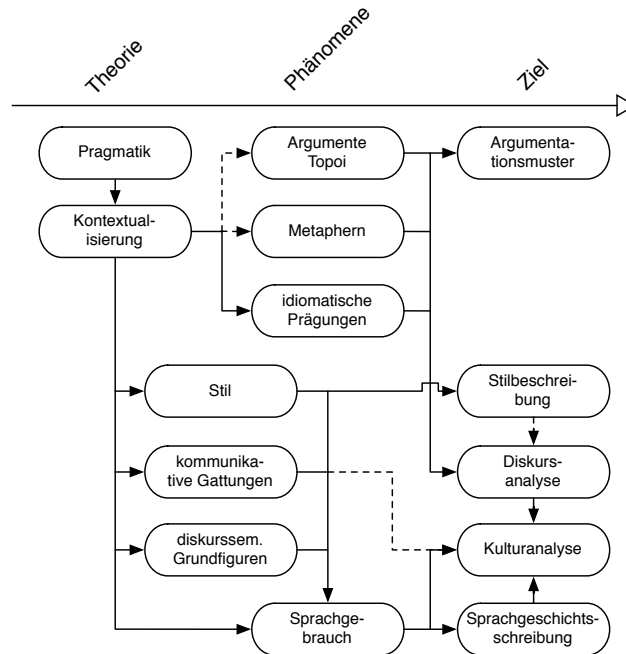
Eine quantitative Methode bietet auch den Vorteil, Sprachgeschichte nicht nur im großen Maßstab als Geschichte von Jahrzehnten oder Jahrhunderten zu betreiben, sondern auch kleinere Zeiträume in den Blick nehmen zu können, da die Veränderungen im Sprachgebrauch empirisch belegt werden können.

In einem abschließenden Kapitel zu dieser Tour d'Horizon über linguistische Konzepte, die den Sprachgebrauch historisch zu fassen versprechen, muss geklärt werden, ob eine solche Methode der kulturalistischen Sprachgebrauchsgeschichtsschreibung die Fragen der vorliegenden Arbeit beantworten kann, und in welchem Verhältnis sie zu den anderen analytischen Konzepten steht.

#### 4.8 Die Einordnung der Konzepte

Die dargestellten Konzepte basieren auf unterschiedlichen theoretischen Annahmen und leisten Unterschiedliches. Die Übersicht in Abbildung 4.3 auf der nächsten Seite versucht diese nach den Kriterien ‚Theorie‘, ‚Phänomen‘ und ‚Ziel‘ zu systematisieren.

Neuere Überlegungen in der Pragmatik, vor allem vor dem Hintergrund der Idee der Kontextualisierungskraft von Sprache, bilden die Grundlage diverser Konzepte und Methoden, die alle neben ähnlichen auch differierende Fragestellungen aufweisen. Interessant ist der Bereich der Phänomene, die unter unterschiedlichen theoretischen Vorbedingungen gefasst werden. Dabei müssen eher konkrete und textoberflächliche Phänomene, wie Topoi, Metaphern oder idiomatische Prägungen, von komplexeren Phänomenen auf der semantischen Tiefenstruktur unterschieden werden. Dazu gehören der Stilbegriff, die diskurssemantischen Grundfiguren und teilweise auch die kommunikativen Gattungen. Für diese tiefenstrukturellen Phänomene stellt



**Abbildung 4.3:** Einordnung von Konzepten zu Sprachgebrauch in die Kategorien ‚Theorien‘, ‚Phänomene‘ und ‚Ziele‘.

sich die Frage nach der Operationalisierung dringender als für die anderen Phänomene.

Alle dargestellten Phänomene lassen sich für Diskurs- und Kulturanalysen nutzen. So wird ein komplexes Ensemble theoretischer Fragmente und Methoden sichtbar, das in den Dienst von Diskurs- und Kulturanalysen gestellt werden kann, so lange dieses Ensemble den Fokus auf den Sprachgebrauch richtet.

Ich versuche nun in einem zweiten Schritt die referierten Konzepte genauer zu charakterisieren und vergleichbar zu machen. Die Abbildung 4.4 auf Seite 90 positioniert die Konzepte auf vier Achsen. Diese Achsen ergeben sich aus den folgenden Polen:

1. Musterhaft vs. singulär: Beruht das Konzept eher auf musterhaft auftretenden oder singulären Phänomenen?

2. Oberflächlich vs. tiefensemantisch: Orientiert sich das Konzept eher an Oberflächenphänomenen oder an tiefensemantischen Kategorien?
3. Gestaltend vs. gestaltet: Schreibt das Konzept dem Phänomen kontextualisierende, Kultur oder Diskurs gestaltende Kraft zu? Oder wird das Phänomen eher als Resultat von Kontexten betrachtet, als von außersprachlichen Faktoren gestaltet?
4. Induktiv vs. deduktiv: Wie muss das Potenzial, das Phänomen induktiv aus Sprachdaten abzuleiten, eingeschätzt werden? Ist das Konzept damit eher durch eine induktive oder deduktive Analyse greifbar?

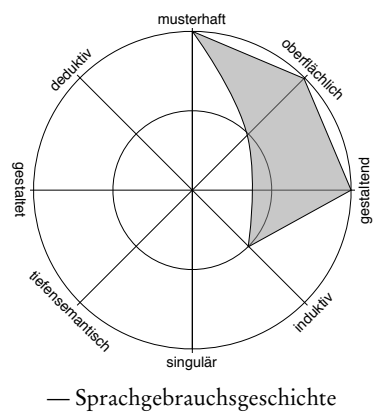
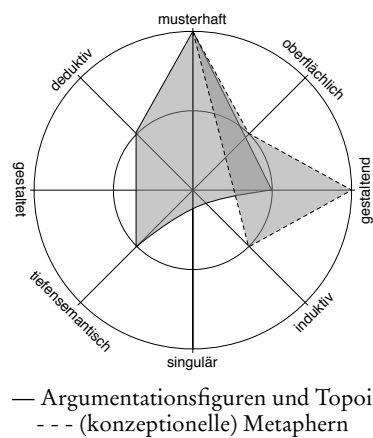
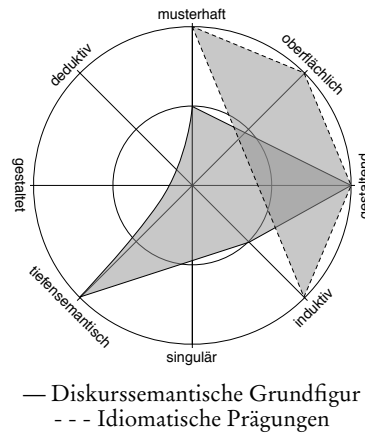
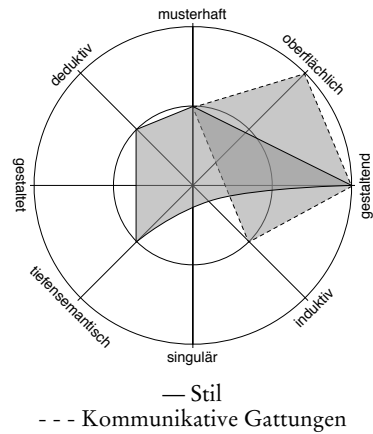
Die Positionierung der Konzepte auf diesen vier Achsen könnte eine objektive Bewertung suggerieren. Dem ist natürlich nicht so. Die Charakterisierung erfolgt deshalb sehr grob auf jeweils zwei Stufen pro Halbachse. Am Beispiel der Achse musterhaft–singulär erläutert: Es werden nur die Stufen stark musterhaft, musterhaft, singulär, stark singulär unterschieden.

Die Spinnenprofile<sup>40</sup> in Abbildung 4.4 auf der nächsten Seite zeigen grobe Einordnungen der Konzepte, die in den obigen Kapiteln diskutiert wurden. Bei aller Ungenauigkeit dieser Einordnungen wird trotzdem deutlich, dass die Konzepte jeweils auffallend unterschiedliche Aspekte von Sprachgebrauch betrachten. Alle Konzepte gehen m. E. von einer gestaltenden, kontextualisierenden Wirkung der Phänomene aus, die sie betrachten, oder es gibt zumindest Ansätze, die die Konzepte so fassen.<sup>41</sup> Unterschiedlich ist die Orientierung auf der Achse zwischen Textoberfläche und semantischer Tiefenstruktur. So sind die idiomatischen Prägungen in erster Linie Phänomene auf der Textoberfläche. Auch die Sprachgebrauchsgeschichte analysiert (zunächst) Phänomene der Textoberfläche.

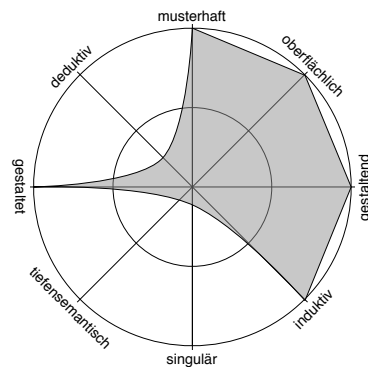
<sup>40</sup> Diese Darstellung in Spinnenprofilen ist von einer Visualisierung aus dem Kontext der politischen Geografie inspiriert: Michael Hermann und Heiri Leuthold, *sotomo*, Universität Zürich.

<sup>41</sup> Das Konzept der Metaphern beispielsweise erscheint in der Forschung in sehr unterschiedlichen Facetten. Die klassische Metaphernlehre geht wohl kaum von einer kontextualisierenden Wirkung aus; die neuere Forschung, wie sie in Kapitel 4.6 dargestellt wurde, unterstellt der konzeptionellen Metapher zumindest teilweise diese Wirkung.





**Abbildung 4.4:** Spinnenprofile semantischer Kategorien und linguistischer Analysekonzepte (der Lesbarkeit willen wurden jeweils maximal zwei Konzepte in ein Profil integriert).



**Abbildung 4.5:** Spinnenprofil der Stärken korpuslinguistischer Ansätze.

Es stellt sich nun die Frage, wo die Stärken eines korpuslinguistischen Ansatzes liegen. Abbildung 4.5 versucht dies darzustellen. Bezüglich dreier der vier Achsen sind die Stärken klar: Die maschinelle Analyse von Korpusdaten funktioniert am besten, wenn stark musterhafte Phänomene, die sich auf der Textoberfläche festmachen lassen, analysiert werden. Dabei ist der mögliche induktive Zugang ein Vorteil dieses quantitativen Verfahrens: Musterhafte Strukturen auf der Textoberfläche lassen sich in Daten zeigen, ohne dass vor Analysebeginn bereits ganz klar festgelegt ist, wonach gesucht werden soll.

Gegenüber der Achse *gestaltend – gestaltet* präferiert die Korpuslinguistik keine bestimmte Position, da es sich hier um eine sehr grundlegende Prämisse handelt. Diese kann sich zwar auf eine Methode auswirken, jedoch nicht solchmaßen, dass sie über die Anwendbarkeit der Methode entscheiden würde. Ein korpuslinguistischer Zugang kann sowohl im Dienste einer Theorie sein, die von der gestaltenden Kraft eines sprachlichen Phänomens ausgeht, als auch einer Theorie dienen, die Sprache mittels außersprachlicher Faktoren gestaltet sieht.

Die Schwächen korpuslinguistischer Zugänge liegen eindeutig darin, komplexe Phänomene zu erfassen, die nicht auf der Textoberfläche festgemacht werden können. Ebenfalls schwierig ist die Analyse singular auftretender Ereignisse, da Korpora keine negative Evidenz bieten: Dass ein Phänomen in den Daten nicht auftritt, bedeutet nicht, dass es nicht existiert. Singularität muss jedoch nicht mit Seltenheit deckungs-

gleich sein, sofern Seltenheit musterhaft ist: Dass ein Phänomen in bestimmten Teilkorpora seltener auftritt als in Vergleichskorpora kann sehr wohl quantitativ-statistisch berechnet werden.

Vergleicht man nun das Spinnenprofil der Korpuslinguistik (Abbildung 4.5 auf der vorherigen Seite) mit den Profilen der restlichen Kategorien (Abbildung 4.4 auf Seite 90), wird klar, welche dieser Kategorien am besten mit korpuslinguistischen Methoden behandelt werden können. Es sind dies die idiomatischen Prägungen und die Ansätze der Sprachgebrauchsgeschichte. Teilweise ergeben sich auch Ähnlichkeiten zu den Konzepten Stil und Metaphern. Es ist jedoch offensichtlich, dass mit korpuslinguistischen Mitteln allein nicht alles erfasst werden kann, was als Element von Stil oder als Metaphern erfasst würde.

In meinen theoretischen Überlegungen zu den Fragen nach den Sprechweisen im Lichte einer Diskursanalyse versuchte ich Folgendes aufzuzeigen:

1. Der Musterbegriff ist zentral, um typische Sprechweisen zu fassen.
2. Es gibt eine Reihe semantischer Kategorien, die fruchtbar gemacht werden können, um Sprechweisen in Diskursen zu untersuchen. Allerdings eignen sich nicht alle gleich gut für einen korpuslinguistischen Zugang, der auf Phänomene auf der Textoberfläche angewiesen ist.
3. Die Konzepte der idiomatischen Prägung und der Sprachgebrauchsgeschichte legen jedoch nahe, dass die Beschränkung auf die Textoberfläche keine Einschränkung darstellt. Die Textoberfläche muss Ausgangspunkt für diskurs- oder kulturanalytische Fragestellungen sein. Genauer: Die Musterhaftigkeit der Textoberfläche. Diese Musterhaftigkeit wird damit zum verbindenden Element zwischen dem Konzept der idiomatischen Prägung und dem theoretischen Ansatz der Sprachgebrauchsgeschichte: Letztere beobachtet die Verwendung und die Funktion von idiomatischen Prägungen, von Mustern.

Im nächsten Teil werde ich nun eine Methode für eine korpuslinguistische Diskursanalyse entwickeln, die auf diesen Überlegungen fußt.



## **II**

### **Mustererkennung als Basis für die Analyse: Methodische Herleitung**



Im ersten Teil der vorliegenden Arbeit wurde das theoretische Fundament für eine korpuslinguistische Diskursanalyse gelegt. Dabei wurde deutlich, dass über musterhafte Strukturen des Sprechens die Gestalt, Funktionsweise und die Grenzen eines Diskurses beschrieben werden können. Der zweite Teil leistet nun die Operationalisierung der Ansätze, um typische Sprechweisen mit korpuslinguistischen Mitteln finden zu können.

Die maschinelle Korpuslinguistik, die mit großen Datenmengen arbeitet, nutzt die lange Tradition der computerlinguistischen Sprachverarbeitung. Die Computerlinguistik entwickelte eine vielfältige Palette an Methoden und Werkzeugen. Diese versuche ich in den Dienst einer korpuslinguistischen Diskursanalyse zu stellen. Daneben sind auch statistische Hilfsmittel nützlich, die in einem eigenen Kapitel diskutiert werden.

Die Korpuslinguistik ist aber weit mehr als ein Methodenapparat und Werkzeugkasten. Ihre Art der Sprachanalyse führte zu wichtigen Paradigmenwechseln in der linguistischen Theorie: Stichwort ist hier das Prinzip einer ‚corpus-‘ oder ‚data-driven‘ Korpuslinguistik, die mitunter traditionelle Sichtweisen auf Sprache auf den Kopf stellt. Dieser Paradigmenwechsel kommt einer linguistischen Diskursanalyse sehr entgegen und ist die korpuslinguistische Antwort auf das theoretische Bedürfnis, wie ich es im ersten Teil formuliert habe.

In den ‚Grundsätzen‘ (Kapitel 5) werde ich erstens auf genau diesen theoretischen Ansatz der Korpuslinguistik eingehen und aufzeigen, wie daraus eine Heuristik entwickelt werden kann, um musterhafte Strukturen in Textkorpora sichtbar zu machen und beschreiben zu können. Zweitens muss diese Heuristik operationalisiert werden: Zunächst geht es darum, mit korpuslinguistischen Mitteln die ‚Klumpen im Text‘ zu definieren und zu finden (Kapitel 6). Damit wird das Feld der Statistik betreten: Auch darin lohnt sich eine Auseinandersetzung mit den verfügbaren Methoden, die nicht einfach Hilfsmittel sind, sondern auf theoretischen Konzepten fußen, die mit meinen kompatibel sein müssen (Kapitel 7). Danach folgt jener Schritt der Heuristik, der als Scharnier zwischen ‚corpus-driven‘ und ‚corpus-based‘ dient



und dessen Ziel in der Beschreibung eines oder mehrerer Diskurse liegt (Kapitel 8). Eine kompakte Zusammenfassung der im Folgenden entwickelten Heuristik kann auch in Bubenhofer (2008a) nachgelesen werden.

Abgerundet wird Teil II mit einer Prüfung bestehender Ressourcen der Korpus- und Computerlinguistik, die für die Beispielanalysen in Teil III verwendet werden (Kapitel 9).

## 5 Grundsätze

Grundsätzliche methodische Überlegungen stehen am Anfang der Diskussion über die Operationalisierung einer korpuslinguistischen Diskursanalyse. Drei Aspekte müssen dabei in den Fokus gerückt werden.

1. Welchen Stellenwert hat das Korpus für die Untersuchung? Da ich davon ausgehe, dass sich Sprachgebrauch in musterhaften Strukturen niederschlägt, muss das Korpus derart befragt werden, dass sich diese Strukturen zeigen. Doch wie kann ich ein Korpus befragen, wenn ich nicht weiß, in welchen Strukturen sich Sprachgebrauch niederschlägt?  
Die Korpuslinguistik hat mit der corpus-driven-Perspektive eine Perspektive eingenommen, die genau eine solche Korpusbefragung ermöglicht. In Kapitel 5.1 stelle ich diese Perspektive vor.
2. Welcher Weg muss gewählt werden, um von Korpusdaten zu einer Beschreibung von Sprachgebrauch und Diskurs zu gelangen? Kapitel 5.2 schlägt eine Heuristik vor, die das leisten soll.
3. Eine Untersuchung braucht eine klar definierte Datenbasis. Worin besteht sie in der corpus-driven Korpuslinguistik? Und wie harmonisiert diese Definition mit der Vorstellung eines Foucault'schen Diskursbegriffs? Diese Fragen müssen in Kapitel 5.3 geklärt werden.

### 5.1 Daten lesen: ‚corpus-based‘ und ‚corpus-driven‘

Sind Korpora nur Belegsammlungen oder Zettelkästen in elektronischer Form? Mitnichten! In entsprechender Größe [...] und mit den entsprechenden Analysemethoden eröffnen sie eine eigene Perspektive in der linguis-

tischen Forschung – die korpuslinguistische Perspektive. (Perkuhn/Belica 2006, 2)

Das Zitat verweist auf ein in der Linguistik noch immer häufiges Missverständnis von Korpuslinguistik. Natürlich können Korpora auch als Nachschlagewerke benutzt werden, um zu überprüfen, ob sich darin ein bestimmtes sprachliches Phänomen belegt findet. Doch gerade mit Hilfe der immer schneller werdenden Computer und der Verfügbarkeit von elektronischen Korpora ist für viele Fragestellungen eine andere Perspektive in der Korpuslinguistik interessanter: die ‚corpus-driven‘-Perspektive:

While corpus linguistics may make use of the categories of traditional linguistics, it does not take them for granted. It is the discourse itself, and not a language-external taxonomy of linguistic entities, which will have to provide the categories and classifications that are needed to answer a given research question. This is the corpus-driven approach. (Teubert 2005, 4)

Dieser Zugang zeichnet sich also im Versuch aus, das Korpus als Datenbestand aufzufassen, in dem mit geeigneten Methoden Strukturen sichtbar gemacht werden, die erst im Nachhinein klassifiziert werden.

Damit steht dieser Zugang in Kontrast zur ‚corpus-based‘-Perspektive, bei der mit bestimmten Kategorien und bestehenden Theorien das Korpus analysiert wird und das Interesse dann darin besteht, eine Hypothese zu testen.<sup>1</sup> Letztlich wird corpus-based die Frage verfolgt, ob das gesuchte Phänomen im Korpus auftritt, wenn ja, wo, wie oft und wie:<sup>2</sup>

[...] the term corpus-based is used to refer to a methodology that avails itself of the corpus mainly to expound, test or exemplify theories and de-

<sup>1</sup> Ich verwende im Folgenden die englischen Termini ‚corpus-based‘ und ‚corpus-driven‘, obwohl in der Literatur teilweise deutsche Übersetzungen zu finden sind. ‚Corpus-based‘ kann zwar gut mit ‚korpusbasiert‘ übersetzt werden, die Übersetzung von ‚corpus-driven‘ mit ‚korpusgesteuert‘ finde ich unpassend, da es sich weniger um ein ‚Steuern‘ handelt, sondern um einen ‚Antrieb‘. Die Übersetzung mit ‚korpusgetrieben‘ klingt jedoch eher ungewohnt.

<sup>2</sup> Neben diesen beiden Perspektiven findet sich bei Tummers u. a. (2005) eine weitere (wobei dort ‚corpus-driven‘ nicht erwähnt wird): ‚corpus-illustrated linguistics‘, die Korpora nicht systematisch und quantitativ auswerten, sondern bloß anekdotisch. Atkins u. a. (1992, 14) erwähnen die beiden Lager „knowledge based“ und „self organizing“ in der Computerlinguistik. Letzteres versucht „to use the statistical regularities to be found in mass text as a key to analysing and processing it“ während ersteres „brings in the results of linguistic theory and logic as the foundation for its models of language“. Vgl. dazu auch Tognini-Bonelli (2001, 66).

scriptions that were formulated before large corpora became available to inform language study. (Tognini-Bonelli 2001, 65)

Doch Korpora mit ganz bestimmten Theorien als Prämissen zu befragen, birgt die Gefahr, in den Daten nur die Strukturen zu finden, die mit der Theorie kompatibel sind und blind gegenüber Evidenzen zu sein, die quer zu einer Theorie stehen:

[...] corpus-based linguists adopt a „confident“ stand with respect to the relationship between theory and data in that they bring with them models of language and descriptions which they believe to be fundamentally adequate, they perceive and analyse the corpus through these categories and sieve the data accordingly. (Tognini-Bonelli 2001, 66)

So sind im corpus-based-Verfahren Korpora nützlich, um aufgrund der Daten die Hypothesen zu korrigieren und zu quantifizieren. Aber:

In this case, however, corpus evidence is brought in as an extra bonus rather than as a determining factor with respect to the analysis, which is still carried out according to pre-existing categories; although it is used to refine such categories, it is never really in a position to challenge them as there is no claim made that they arise directly from the data. (Tognini-Bonelli 2001, 66)

So ist es oft naheliegend, Daten, die einem Modell widersprechen, als durch Sprachvarietät verursachte Ausnahmen zu erklären, die jedoch zu marginal sind, als dass sie das Modell angreifen könnten.<sup>3</sup> Die corpus-driven-Analyse bietet einen anderen Zugang, der die Daten zum Ausgangspunkt der Theoriebildung macht.

Das corpus-driven-Paradigma ist nicht neu. Bei Sinclair (1991) bereits angedacht, wird es bei Tognini-Bonelli (2001) mit dem Terminus „corpus-driven Linguistics“ (CDL) explizit gemacht. Doch sind es vor allem im deutschsprachigen Raum auch heute erst wenige Linguistinnen und Linguisten, die konsequent so arbeiten. Dazu gehören z. B. Arbeiten von Kathrin Steyer (Steyer 2004a, Steyer 2004b, Steyer/Lauer 2007, Steyer/Brunner 2009), die auf Konzepten und Methoden von Cyril Belica und Rainer Perkuhn fußen (Belica 1996, Belica 2001–2006, Perkuhn u. a. 2005, Perkuhn/Belica 2006). Diese Methoden können im

<sup>3</sup> Tognini-Bonelli (2001, 67) spitzt dieses Dilemma zur rhetorischen Frage zu: „A problem then arises, one that has been cropping up throughout the history of linguistics: given that the data is non-negotiable, does the linguist choose to revise the theory and derive it more directly from corpus evidence, or does (s)he opt to insulate the data from the theory?“.

Rahmen des Korpus des IDS (DeReKo IDS o. J.) teilweise angewendet werden, um corpus-driven arbeiten zu können.<sup>4</sup>

Es ist unumgänglich corpus-driven zu beginnen, um musterhafte Strukturen in Korpora zu finden, die der Diskursanalyse dienen sollen.<sup>5</sup> Denn damit scheint es am ehesten möglich, zum einen Strukturen aufzudecken, die ihre Wirkung im Diskurs mehr oder weniger verdeckt entfalten und gleichzeitig aufgrund empirischer Evidenz Kategorien zu bilden, die nicht unbedingt mit bestehenden (linguistischen) Kategoriensystemen übereinstimmen müssen.

So nennen auch Steyer/Lauer (2007, 493) die folgenden empirischen Grundprinzipien als aufeinander aufbauende Schritte:

- Beobachtung der Sprachdaten;
- Akzeptanz aller Evidenzen;
- Bildung von Hypothesen auf Basis der Evidenzen;
- empirisches Prüfen der Hypothesen (Zusammenspiel von Induktion und Deduktion);
- deskriptive Aussagen, die die Evidenzen reflektieren;
- Generalisation im Sinne von Gebrauchsregeln.

(Steyer/Lauer 2007, 493)

Bei diesen Grundprinzipien wird aber gleichzeitig sichtbar, dass auch die corpus-driven-Sicht alleine nicht ausreicht. Mit dem „Zusammenspiel von Induktion und Deduktion“ wird die corpus-based-Perspektive wieder ins Spiel gebracht. Auch ich vertrete die Auffassung, dass dies nötig ist; im nächsten Kapitel soll dies dargelegt werden.

## 5.2 Daten, Muster, Beschreibung: Die Heuristik im Überblick

Ziel der hier zu entwickelnden Heuristik ist es, aus Daten musterhafte Strukturen zu gewinnen, diese zu interpretieren und damit eine Diskursbeschreibung zu erstellen.

Die Heuristik umfasst folgende Schritte:

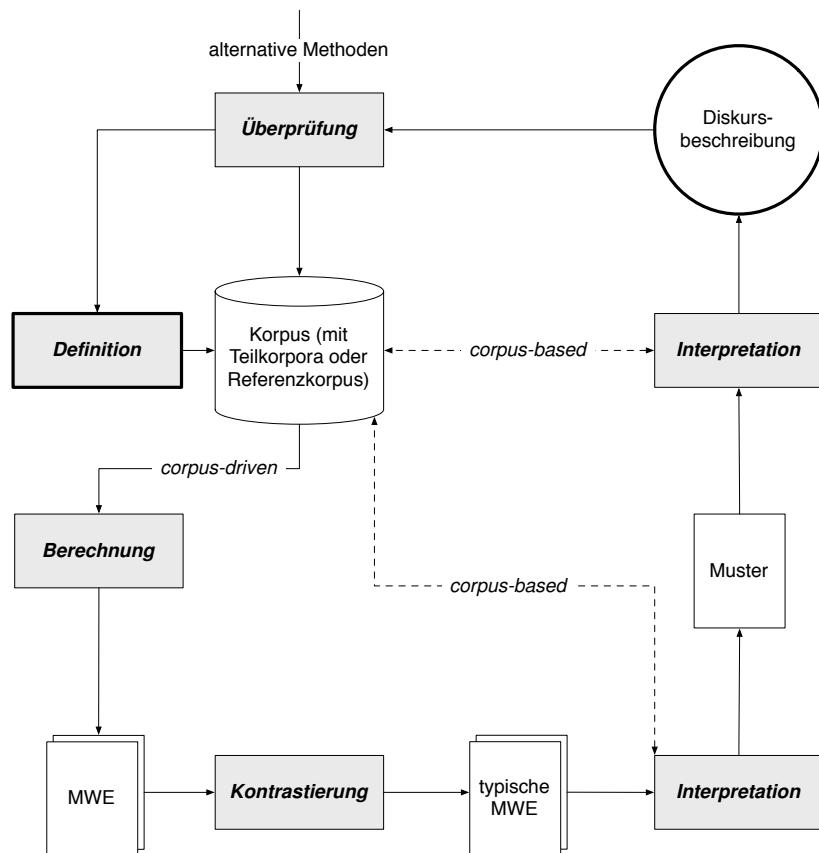
<sup>4</sup> Zu den Details dazu später in Kapitel 6 und 9.

<sup>5</sup> Vgl. dazu auch das Plädoyer von Glasze (2007) und Glasze/Mattisek (2009), den Ausgangspunkt bei lexikometrischen Analysen zu setzen, um diskursanalytisch zu arbeiten.

1. Es wird ein Korpus in Kombination mit einem oder mehreren Referenzkorpora definiert, wobei auch Teile des Untersuchungskorpus als Referenzkorpus dienen können.
2. Aus dem Korpus und den Referenzkorpora werden corpus-driven Listen von Mehrworteinheiten berechnet.
3. Durch Kontrastierungen der Listen untereinander, können die für bestimmte Teilkorpora typischen Mehrworteinheiten berechnet werden.
4. Nun erfolgt unter corpus-based-Rückgriffen ins Korpus die Interpretation der Mehrworteinheiten, um aus ihnen abstraktere Sprachgebrauchsmuster abzuleiten.
5. Die weitere Analyse der Verwendung dieser Sprachgebrauchsmuster – ebenfalls corpus-based – führt zu einer Diskursbeschreibung.
6. Die Diskursbeschreibung muss aufgrund der Korpusdaten auf ihre Plausibilität hin geprüft werden. Dabei ist es sinnvoll, die Daten auch mit alternativen Methoden auszuwerten, um die Diskursbeschreibung einer erweiterten Prüfung zu unterziehen. Je nach Resultat dieser Überprüfungen muss der Prozess mit veränderten Parametern erneut durchlaufen werden, um zu einer korrigierten Diskursbeschreibung zu gelangen.

Die Abbildung 5.1 auf der nächsten Seite zeigt den Vorgang im Überblick. Dabei wird die Heuristik deutlich, die das Korpus unter den zwei Paradigmen corpus-driven und corpus-based nutzt: Im ersten Schritt werden aus dem Korpus Kandidaten für musterhafte Strukturen gewonnen, ohne den Fokus bereits auf bestimmte Muster einzuschränken. Nachgelagerte Prozesse der Kontrastierung der Muster gruppieren diese nach unterschiedlichen Kriterien entweder auf diachroner oder synchroner Achse. Dadurch entsteht eine leichter handhabbare Menge an Mustern, die interpretiert werden kann.

Der interpretative Schritt stellt das Scharnier zwischen den beiden Perspektiven corpus-driven und corpus-based dar. Neben dem Wechsel zwischen corpus-driven und corpus-based bewegt sich die Heuristik auch zwischen quantitativen und qualitativen Methoden.



**Abbildung 5.1:** Die Operationalisierung einer korpuslinguistischen Diskursanalyse im Überblick.

Die Entscheidung darüber, welche der berechneten Mehrwort-einheiten weiter verfolgt werden, wie die Belege im Korpus gewichtet werden, um die Sprachgebrauchsmuster abzuleiten und letztlich wie die weitere Abstrahierung der Befunde zu einer Diskursbeschreibung erfolgt, sind qualitativ-interpretative Akte. Allerdings beruhen sie auf einer empirischen Basis.

Wie bereits oben erwähnt, verläuft der Prozess zirkulär, da die Beschreibung laufend überprüft wird und der Prozess mit Berechnungen,

z. B. mittels veränderter Parameter, neu gestartet werden kann. Neue Interpretationen auf Basis des Korpus führen so zu Korrekturen der Beschreibung. Es kann auch notwendig werden, die Definition des Korpus aufgrund neuer Fragestellungen, die sich aus der Diskursbeschreibung ergeben, zu verändern.

### 5.3 Wörter, Sätze, Texte? Die Elemente und Grenzen eines Diskurses bestimmen

Beim korpuslinguistischen Zugang zu Diskursen stellt sich ein praktisches Problem: Das zu untersuchende Korpus muss aus einer endlichen Menge von textuellen Einheiten bestehen. Die Definition von ‚Korpus‘ als Datenbestand sagt noch nichts darüber hinaus, welche textuellen Einheiten diesen Bestand ausmachen. Nahe liegend ist es, das Korpus als eine Menge von ausgewählten Texten zu definieren. Doch ebenso möglich ist es, Absätze oder Sätze als kleinste Einheiten zu verwenden, oder aber eine andere Größe wie ‚Aussagen‘, wie immer diese Größe auch definiert wird. Das hieße, dass aus den Einheiten im Korpus nicht zwingend Texte rekonstruierbar sind, sondern dass diese unter Umständen nur zu Teilen im Korpus repräsentiert sind.

Diese Möglichkeit kommt der Foucault’schen Auffassung von Diskurs entgegen: Die Diskursanalyse möchte „Spiele von Beziehungen“ beschreiben, und dabei alle

Beziehungen der Aussagen untereinander (selbst wenn diese Beziehungen dem Bewusstsein des Autors entgehen; selbst wenn es sich um Aussagen handelt, die nicht den gleichen Autor haben; selbst wenn diese Autoren einander nicht kennen); Beziehungen zwischen so aufgestellten Gruppen von Aussagen (selbst wenn diese Gruppen nicht die gleichen Gebiete oder benachbarte Gebiete treffen; selbst wenn sie nicht das gleiche formale Niveau haben; selbst wenn sie nicht der Ort bestimmbarer Austausches sind); Beziehungen zwischen Aussagen oder Gruppen von Aussagen oder Ereignissen einer ganz anderen (technischen, ökonomischen, sozialen, politischen) Ordnung (Foucault 1981, 45)

analysieren. Dabei sollen die diskursiven Einheiten kontrolliert in „andere Einheiten“ gruppiert werden, Mengen, die „nicht arbiträr [...], indessen aber unsichtbar geblieben wären“ (Foucault 1981, 45).

Dies bedeutet:



1. Es werden Aussagen analysiert, von denen angenommen werden muss, dass sie nicht nur in Form von abgeschlossenen Texten, sondern auch in anderen textuellen Größen auffindbar sind.
2. Die „Spiele von Beziehungen“ machen nicht an den Grenzen von eindimensional (z. B. thematisch) definierten textuellen Einheiten Halt.

Für diese Ausweitung des Untersuchungsobjekts der Diskursanalyse plädiert auch Jung (1996) und weitet so die Definition von ‚Diskurs‘ von Busse u. a. (1994, 14) aus:

Meines Erachtens beziehen sich [...] nicht ganze Texte aufeinander, sondern Aussagen, Behauptungen, Topoi – über den angemessensten Begriff bin ich mir nicht im klaren, da alle Ausdrücke schon auf die eine oder andere Weise besetzt sind. (Jung 1996, 460)

So bestehen Texte aus Aussagen, die aber unterschiedlichen Diskursen angehören können, so dass auch ein einzelner Text gleichzeitig unterschiedlichen Diskursen angehört.

Allerdings sieht Jung (1996, 463) forschungspraktische Schwierigkeiten, dieser Definition von Diskurs konsequent zu folgen; der Text bleibe letztlich doch die Orientierungsgröße bei der Korpuszusammenstellung. Dies scheint mir jedoch in erster Linie ein technisches, und damit prinzipiell lösbares Problem zu sein, was auch Spitzmüller (2005, 46f.) zeigt. Er arbeitet mit einem Korpus aus Textfragmenten, die in einer Datenbank verwaltet werden. Damit ist es möglich, daraus Teilkorpora als bestimmte Zusammenstellungen von Textfragmenten zu bilden, als auch jederzeit die Einheit Text, in der eine bestimmte Aussage erscheint, rekonstruieren zu können. Spitzmüller (2005, 44ff.) stellt deutlich dar, wie üblich Diskursüberlagerungen in Texten sind und dass die Analyse unvollständig wäre, würden diese Überlagerungen nicht berücksichtigt. In seinem Korpus zur Untersuchung des Anglizismendiskurses macht er in einem Viertel der 921 Dokumente mit Hauptdiskurs ‚Anglizismen‘ andere metasprachliche Diskurse aus. Und das gesamte Korpus aus 1380 Dokumenten enthält 459 Dokumente, die Anglizismen nur als Nebendiskurs führen, in den Augen von Spitzmüller (2005, 78) aber wichtig sind, da sich an ihnen ablesen lässt, „inwieweit der Diskurs über den eigentlichen metasprachlichen Gesamtdiskurs hinaus wirkt“. Ohne Fragmentierung der Texte wäre eine so differenzierte Analyse nur schwer möglich.

Die Fragmentierung der Texte löst allerdings noch nicht das Problem der Korpuszusammenstellung. Wie wird die Menge der Textfragmente definiert, die untersucht werden soll?

Ich versuche dieses Problem corpus-driven anzugehen. Damit wird klar, dass das Korpus zunächst gar nicht – oder etwas genauer: nur sehr unpräzise – definiert wird.<sup>6</sup> Die Chance, corpus-driven musterhafte Strukturen zu finden, die hinsichtlich einer Forschungsfrage von Relevanz sind, erhöht sich, wenn das Korpus zunächst möglichst offen gestaltet ist. Oder anders ausgedrückt: Erst die Analyse der musterhaften Strukturen definiert die Grenzen des Korpus, das eine Menge ähnlicher diskursiver Einheiten enthält.<sup>7</sup> Das bedeutet: Dieselben methodischen Prinzipien, wie sie für die Berechnung der musterhaften Strukturen verwendet werden, schlage ich auch für die Eingrenzung des Korpus vor. In Kapitel 6 werde ich detailliert auf Methoden der Berechnung von Mustern auf Wortebene eingehen. Daneben sind zusätzlich andere Clusterberechnungen denkbar, die weiter unten kurz Erwähnung finden sollen.

6 Ein ähnliches Vorgehen, nämlich die Arbeit mit einem „offenen Korpus“, schlagen auch Busse/Teubert (1994, 17) vor. Ansatzweise induktiv arbeitet auch Wengeler (2003, 296f.), der Topoi des Migrationsdiskurses quantitativ untersucht und inhaltsanalytisch mit einem Kategoriensystem arbeitet und erst „nach einer ersten Lektüre (eines Teils) des Textmaterials die Kategorien definiert, die dazu führen sollten, die Forschungsfrage‘ zu beantworten, d. h. es wurden die kontextspezifischen Topoi definiert, die in den Texten vorkamen und die bei der Auswertung dieser und weiterer Texte aufgesucht und gezählt wurden“.

7 Biber (1994, 399f.) (vgl. auch Biber 1993) bestätigt die Schwierigkeit, ein bezüglich einer definierten Grundgesamtheit repräsentatives Korpus zu erstellen. Möchte man eine geschichtete Zufallsstichprobe vornehmen und werden die Schichten nicht nach außerlinguistischen Faktoren wie demographischen Eigenschaften von Textproduzentinnen und -rezipienten definiert, sondern nach textinternen Kriterien wie Textsorte, ist es zunächst unmöglich, die Stichprobenmengen pro Schicht festzulegen. Denn diese Menge müsste proportional zum Umfang der Schichten in der Grundgesamtheit stehen. Der Umfang dieser linguistisch definierten Schichten in der Grundgesamtheit ist aber unbekannt. In einem zirkulären Verfahren über Pretests muss zuerst der Umfang dieser Schichten in unterschiedlichen Stichproben abgeschätzt werden, damit die Größe der Stichproben proportional zu der in der Grundgesamtheit erwarteten Verteilung festgelegt werden kann: „Regardless of the initial design, the compilation of a representative corpus should proceed in a cyclical fashion: a pilot corpus should be compiled first, representing a relatively broad range of variation but also representing depth in some registers and texts. [...] Then empirical research should be carried out on this pilot corpus to confirm or modify the various design parameters. Parts of this cycle could be carried out in an almost continuous fashion, with new texts being analyzed as they become available, but there should also be discrete stages of extensive empirical investigation and revision of the corpus design“ (Biber 1994, 400).

Es ist jedoch klar, dass der Forderung nach einem möglichst breiten Korpus, das erst im Lauf der Untersuchung eingegrenzt wird, in der Forschungspraxis nicht entsprochen werden kann. Denn aus einem tendenziell unendlich großen Korpus würden auch tendenziell unendlich viele Typen von musterhaften Strukturen und damit Diskursen hervorgehen. Es muss also eine Grundgesamtheit definiert werden, die diese Zahl von Diskursen auf eine Textmenge reduziert, die potenziell eine dem Forschungsinteresse entsprechende Auswahl von Diskursen zeigt. Dabei muss jedoch darauf geachtet werden, dass die Reduzierung eine thematische und/oder textsortenspezifische Varianz bewahrt. Aus dieser Grundgesamtheit schließlich wird eine Stichprobe gezogen, die das Korpus für den Ausgangspunkt bildet.<sup>8</sup> Die Anwendung der in Kapitel 5.2 skizzierten Heuristik erlaubt anschließend immer feinere Grenzziehungen in der Datenmenge der Stichprobe. Diese Grenzziehungen werden dann auf die Grundgesamtheit interpoliert.

Die Heuristik kann sich an unterschiedlichen Dimensionen orientieren, wobei eine thematische nur eine der möglichen Dimensionen ist. Dies zeigt die vorliegende Arbeit: Sie interessiert sich für die Sprechweisen in Diskursen, weniger für thematische Inhalte. Hier wird offensichtlich, dass man sich mit einer thematischen Auswahl von Zeitungstexten bereits unnötig einschränken würde. Die diskursiv interessanten Grenzen verlaufen hier kaum den Themenrändern entlang. Doch selbst wenn das Interesse „Aussagen, Behauptungen und Topoi“ (Jung 1996, 460) zu einem bestimmten Thema gilt, ist nicht zu erwarten, dass diese nur oder hauptsächlich in Texten vorkommen, die primär dieses Thema behandeln und entsprechende Schlagworte enthalten.

Konkret gestaltet sich der Vorgang der Korpusauswahl am Beispiel der vorliegenden Arbeit so: Das Forschungsinteresse liegt im Sprachgebrauch in Diskursen, wie sie in Tageszeitungen wirken. Dies verbunden mit der These, dass diese Sprechweisen sich verändern und dass die diachronen Veränderungen Rückschlüsse auf das Wirken der Diskurse zulassen.

So breit das Forschungsinteresse ist, so rasch wird dieses durch forschungspraktische Zwänge eingeschränkt: Ich beschränke mich

---

<sup>8</sup> Zu Grundgesamtheit und Stichproben: Tognini-Bonelli (2001, 59) und Biber (1994, 378).

im Rahmen der vorliegenden Arbeit auf nur eine Zeitung, die ‚Neue Zürcher Zeitung‘ (‚NZZ‘), wobei andere Zeitungen in den corpus-based-Perspektiven punktuell hinzugezogen werden.<sup>9</sup> Die NZZ ist in elektronischer Form ab 1993 verfügbar. Ich beschränke mich auf den Zeitraum von 1995 bis 2005. Aus diesem Zeitraum ziehe ich eine Zufallsstichprobe von knapp 45 000 Artikeln aus allen Ressorts.<sup>10</sup> Es wäre technisch zwar möglich, wenn auch aufwändiger, mit allen Artikeln des Zeitraums zu arbeiten. Die Vereinbarung mit dem Verlag erlaubt dies jedoch nicht.

‚Zufallsstichprobe‘ bedeutet, dass jedes zu befragende Element der Grundgesamtheit (also jeder Artikel) mit gleicher Wahrscheinlichkeit  $> 0$  Teil der Stichprobe werden kann.<sup>11</sup> Damit kann diese Stichprobe als repräsentativ für die Grundgesamtheit aller NZZ-Artikel zwischen 1995 und 2005 angesehen werden.<sup>12</sup> Es ist jedoch klar, dass diese Grundgesamtheit nicht dem ‚allgemeinen Sprachgebrauch‘ entspricht. Alle Aussagen, die aufgrund der Daten gemacht werden können, dürfen statistisch gesehen nur auf den Sprachgebrauch der NZZ im untersuchten Zeitraum bezogen werden.

Diesen Einschränkungen der Repräsentativität der Stichprobe zu einer Grundgesamtheit steht die Offenheit gegenüber, die bezüglich der Artikelauswahl eingenommen werden kann: Verwendet werden die vollständigen Ausgaben (allerdings ohne Werbung, Inserate, Todesanzeigen und Bilder sowie gestalterischen Elementen; auch dies aus forschungspraktischen Gründen) und alle Ressorts werden berücksichtigt. Erst bei der weiteren Analyse wird sich zeigen, wo hier

<sup>9</sup> An dieser Stelle bedanke ich mich für die Erlaubnis des NZZ Medien Verlags, diese Daten für die vorliegende Untersuchung nutzen zu können.

<sup>10</sup> Ich komme in Teil III auf die Eckdaten des Korpus zurück. Die detaillierten Angaben sind deshalb dort in Kapitel 10.2 und insbesondere in Tabelle 10.2 auf Seite 191 zu finden.

<sup>11</sup> Vgl. dazu im linguistischen Kontext Albert/Koster (2002, 28), speziell im korpuslinguistischen Kontext Oakes (1998, 9) und Biber (1994).

<sup>12</sup> Die Statistik kennt die geschichtete Stichprobe (‚stratified sample‘) als mögliche Einschränkung einer einfachen Zufallsstichprobe. Dazu werden zunächst Schichten, z. B. nach demographischen Kriterien, definiert und daraus Stichproben gezogen. Der Umfang der Stichprobe kann sich nach dem Umfang der Schichten in der Grundgesamtheit richten. Biber (1994) schlägt für linguistische Untersuchungen Schichten vor, die sich an der Textproduktion, Textrezeption oder an den Texten selber (z. B. Textsorten) orientieren. Unter Umständen wäre es auch für die vorliegende Untersuchung sinnvoll, eine geschichtete Stichprobe vorzunehmen und z. B. nur bestimmte Ressorts zu berücksichtigen. Aus technischen Gründen war dies jedoch nicht möglich, da die automatisierte Abfrage des NZZ-Korpus nicht nach Ressorts gesteuert werden konnte.

die Grenzen verschiedener Diskurse verlaufen werden. Sie müssen sich prinzipiell nicht mit Ressort- oder Artikelgrenzen decken.

## 6 Musterhafte Strukturen finden: Die Klumpen im Text

Musterhafte Strukturen zeigen sich als rekurrente sprachliche Einheiten: als Klumpen im Text. Der britische Kontextualismus entwickelte mit dem Konzept der Kollokation ein interessantes Mittel, um musterhafte Strukturen in Texten aufzuspüren. Dieser Ansatz ist eine Antwort auf die Linguistik Chomskyscher Prägung, die sich für die Systematik der *Langue* interessiert und dabei die *Parole* völlig vernachlässigt. Der Kontextualismus kehrt dieses Verhältnis zu Gunsten des Sprachgebrauchs um: *Parole* ist die Datenbasis, die analysiert werden muss, um daraus eine Systematik abzuleiten.

Mit dem Kollokationenbegriff wird dieses Prinzip ernst genommen, wie sich beispielsweise in der Semantik zeigt: Die Bedeutung eines Wortes liegt in seinem Gebrauch, oder mit Firth gesprochen in den Begleitern, die es aufweist.

Allerdings gibt es viele Möglichkeiten, den Kollokationenbegriff zu operationalisieren. Dieses Kapitel stellt neben Firths ursprünglichem die neueren Konzepte vor, die auf dem Kollokationenbegriff fußen.

### 6.1 Kollokationen, Kookkurrenzen, n-Gramme

Der Begriff der ‚Kollokationen‘ (engl: *collocations*) geht auf John Rupert Firth (1957, 194) zurück.<sup>1</sup> Firth versucht „Modes of Meaning“ zu definieren, wobei einer der Modi der Modus „meaning by ‚collocation‘“ ist:

The following sentences show that part of the meaning of the word *ass* in modern colloquial English can be by collocation:

- (i) An ass like Bagson might easily do that.
- (ii) He is an ass.
- (iii) You silly ass!
- (iv) Don't be an ass!

---

<sup>1</sup> Eine ausführliche Darstellung des Firthschen Kollokations-Konzepts sowie des britischen Kontextualismus generell bietet Lehr (1996, 21ff.; 7ff.).

One of the meanings of *ass* is its habitual collocation with an immediately preceding *you silly*, and with other phrases of address or of personal reference. Even if you said „An ass has been frightfully mauled at the Zoo“, a possible retort would be, „What on earth was he doing?“

There are only limited possibilities of collocation with preceding adjectives, among which the commonest are *silly*, *obstinate*, *stupid*, *awful*, occasionally *egregious*. *Young* is much more frequently found than *old*. The plural form is not very common. (Firth 1957, 194f.)

Unter Kollokationen werden also häufig auftretende Wortverbindungen verstanden, die für eine Sprache oder einen Teilbereich einer Sprache charakteristisch sind. Diese Teilbereiche können z. B. eine bestimmte Textsorte, ein bestimmtes Thema oder Sprachäußerungen einer bestimmten sozialen Gruppe etc. umfassen. Wie nun ‚charakteristisch‘ und ‚häufig auftretend‘ operationalisiert werden kann, ist unklarer, als es scheinen mag. Firth (1957) führte den Begriff ‚Kollokation‘ ein, um die jeweils für einen bestimmten Sprachausschnitt typische Bedeutung eines Wortes zu definieren, z. B. in Gedichten, und letztlich den sprachlichen Stil eines Autors linguistisch zu beschreiben.<sup>2</sup> „This kind of study of the distribution of common words may be classified into general or usual collocations and more restricted technical or personal collocations“ (Firth 1957, 195). Wie aber die usuellen von den nicht-usuellen Kollokationen getrennt werden sollen, wie Firth fordert, bleibt unklar.

<sup>2</sup> Allerdings bleibt die Frage der Operationalisierbarkeit des Firthschen Kollokationenbegriffs unklar. Es gibt keine eindeutige Definition davon und keinen statistischen Nachweis für das, was Firth ‚typisch‘ nennt. Diese Kritik äußerte auch Haugen (1958, 500f.): „[...] while he illustrates his point with a wealth of examples, he does not establish techniques or definitions that would enable others to apply his ideas with confidence. For one thing it never becomes clear just how many modes or levels he wishes to recognize. In his early essay he distinguished five levels called phonetic, lexical, morphological, syntactical, and semantic. In the later essay I have counted six which he calls ‚lower‘ or ‚minor‘: phonetic, phonaesthetic (a term for which he claims the authorship), phonological, prosodic, collocational, grammatical. Beyond these come the ‚categories of the context of situation,‘ which are the only ones that he would call semantic (Haugen 1958, 500f.)“. Die fehlende Operationalisierung der Typik liegt zum Teil natürlich auch an technischen Limiten. Allerdings existiert eine lange Tradition statistischer Stilbeschreibung, die ihre Anfänge mit Augustus de Morgan in den 1850er-Jahren hat und dreißig Jahre später in erste Berechnungen von sprachlichen Merkmalen zur Charakterisierung des Stils in Stücken von Shakespeare mündet (Tuldava 2005, 370f.). Ebenfalls zum Problem der Operationalisierung des Kollokationenbegriffs bei Firth vgl. Lehr (1996, 22).

Halliday (1961, 276f.) leistet wenig später einen Schritt in Richtung Operationalisierung des Firthschen Kollokationen-Konzepts.<sup>3</sup> Er definiert:

Collocation is the syntagmatic association of lexical items, quantifiable, textually, as the probability that there will occur, at n removes (a distance of n lexical items) from an item x, the items a, b, c . . . . Any given item thus enters into a range of collocation, the items with which it is collocated being ranged from more to less probable; and delicacy is increased by the raising of the value of n and by the taking account of the collocation of an item not only with one other but with two, three, or more other items. (Halliday 1961, 276)

Damit ist der Grundstein gelegt, um Kollokationen über Auftretensfrequenzen definieren zu können, wie das in der Korpuslinguistik heute üblich ist. Zusätzlich werden sie gemeinhin als statistisch auffallend, damit als überzufälliges Aufeinandertreffen von Wörtern, gefasst. Allerdings kann man hier bereits ein erstes terminologisches Problem ausmachen, wenn der Begriff der ‚Kollokation‘ mit jenem der ‚Kookkurrenz‘<sup>4</sup> kontrastiert wird. Wenn in der Literatur Kollokationen von Kookkurrenzen differenziert werden, geschieht dies dadurch, dass Kollokationen als Spezialfall der Kookkurrenzen betrachtet werden. Dabei sind bei den Kollokationen interpretierende Schritte im Spiel, die sie von den puren Kookkurrenzen scheiden. Denn diese werden einfach definiert als gemeinsames Auftreten von bestimmten Wörtern in einem definierten Kontext. Darüber hinaus kann das Moment der Gebräuchlichkeit hinzugezogen werden: Statistisch signifikante Kookkurrenzen sind Wortverbindungen, die überzufällig oft in einer

<sup>3</sup> Halliday (1961, 241) geht es darum, den umfassenden Erklärungsanspruch der Generativen Grammatik zurückzustutzen und die beiden traditionelleren Ansätze der „description in modern linguistics: the ‚textual‘ and the non-textual or, for want of a better word, ‚exemplificatory‘“ zu stärken. Im Bereich der Lexik bedeutet das eine Betonung des „more likely“ und „less likely“ anstelle von „possible“ und „impossible“: „but [...] this particular type of likelihood is not accounted for by grammar, at least not by grammar of the delicacy it has yet attained. It is however too often assumed that what cannot be stated *grammatically* cannot be stated *formally*: that what is not grammar is semantics, and here, some would add, linguistics gives up [...]. But the view that the only formal linguistics is grammar might be described as a colourless green idea that sleeps furiously between the sheets of linguistic theory, preventing the bed from being made. What are needed are theoretical categories for the formal description of lexis. [...] It seems that two fundamental categories are needed, which we may call ‚collocation‘ and ‚set‘“ (Halliday 1961, 275f.).

<sup>4</sup> Von engl. ‚cooccurrences‘; auch ‚Kovorkommen‘ (Lemnitzer 1997, 124).



bestimmten Datenbasis auftreten (Lemnitzer/Zinsmeister 2006, 147f.; Steyer 2004a, 96).<sup>5</sup>

Kollokationen hingegen sind eine Untergruppe von Kookkurrenzen, bei denen die gemeinsam auftretenden Wörter als „strukturell interessante“ (Lemnitzer/Zinsmeister 2006, 196) Einheiten interpretiert werden. Die Gründe, weshalb sie als interessant definiert werden, sind unterschiedlich und teilweise abhängig vom Forschungsinteresse. So sind bei Lemnitzer (1997, 122) Kollokationen arbiträre Wortverbindungen. Mit der Arbitrarität wird betont, dass es sich um ein Gebilde handelt, das aus seinen Bestandteilen nicht erklärt werden kann und deshalb Ausdruck einer Präferenz der Parole ist und das Gebilde deshalb sprachstrukturell gesehen auch anders hätte ausfallen können. So wäre *Zähne putzen* eine Kollokation, da sprachstrukturell auch *Zähne waschen* möglich wäre, aber eben nicht gebräuchlich ist. *Heute Abend* dagegen wäre keine Kollokation, sondern eine Kookkurrenz.<sup>6</sup> In anderen Fällen werden syntaktische Einschränkungen auferlegt, indem etwa verlangt wird, dass „die Glieder einer Kollokation in einer syntaktischen Beziehung zueinander stehen, z. B. als Köpfe einer Verbalphrase und einer gleich- oder untergeordneten Nominalphrase oder als Kopf einer Nominalphrase und Kopf einer untergeordneten Adjektivphrase“ (Lemnitzer/Zinsmeister 2006, 148). Damit wäre *heute Abend* als Kollokation aufzufassen, hingegen die zwar statistisch signifikante, jedoch nicht syntaktische Beziehung *in der ... durch* aber nicht.<sup>7</sup>

<sup>5</sup> Es ist offensichtlich, dass für die meisten Untersuchungen das „singuläre Kovorkommen zweier Ereignissen von geringem Interesse“ ist; relevant seien oft nur die „wiederholt in einer Datenbasis vorkommenden Zeichenkettenpaare“ Lemnitzer (1997, 121).

<sup>6</sup> Ähnlich unterscheidet Evert (2005, 17) und fügt hinzu, dass deswegen Kollokationen in einem Wörterbuch als selbständiger Eintrag aufgeführt werden müssten: „A collocation is a word combination whose semantic and/or syntactic properties cannot be fully predicted from those of its components, and which therefore has to be listed in a lexicon.“

<sup>7</sup> Auch in der Phraseologie-Forschung sind Kollokationen durch ein Plus gegenüber Kookkurrenzen definiert. So verwendet Burger (1998, 50f.) den Terminus als Unterbegriff für Phraseologismen und bezeichnet damit den „Bereich der festen Wortverbindungen, die nicht oder nur schwach idiomatisch sind“. Sie werden somit aber als Einheiten aufgefasst, die bereits als feste Wortverbindungen interpretiert worden sind und nicht einfach eine statistische Auffälligkeit darstellen. Ähnlich Feilke (1996, 115ff.), der aber zeigt, dass dieses Plus auch (entgegen Lemnitzers Vorschlag oben) im pragmatischen Wert definiert werden kann. Weiter einschränkend verwendet Hausmann (1985) den Begriff, indem nur Kombinationen von zwei Wörtern betrachtet und ein

Allerdings wird dieses Plus der Kollokationen gegenüber den Kookkurrenzen z. B. bei Berry-Rogghe (1972) auch bloß als statistische Signifikanz definiert. Er definiert im Anschluss an Halliday (1961):

[T]he aim is to compile a list of those syntagmatic items („collocates“) significantly co-occurring with a given lexical item („node“) within a specified linear distance („span“). „Significant collocation“ can be defined in statistical terms as the probability of the item x co-occurring with the items a, b, c, ... being greater than might be expected from pure chance. (Berry-Rogghe 1972, 103)

Manning/Schütze (2002, 151) sehen das Plus der Kollokationen gegenüber den Kookkurrenzen in der „eingeschränkten Kompositionaltät“,<sup>8</sup> also darin, dass die Bedeutung der Kollokation sich nicht aus den Einzelbedeutungen der Wörter ergibt.<sup>9</sup> *Blinder Passagier* wäre nach dieser Definition also eine Kollokation, *heute Abend* allerdings nicht. Für solche Arten von signifikanten Verbindungen zwischen Wörtern schlagen sie die Bezeichnungen „association“ oder eben „co-occurrence“ vor (Manning/Schütze 2002, 185).

Sinclair (2004, 141) scheint keine explizite Unterscheidung zwischen Kollokationen und Kookkurrenzen zu machen, wenn er definiert: „Collocation [...] is the co-occurrence of words with no more than four intervening words [...]“. Allerdings ist für ihn das Konzept der Kollokationen bzw. der Musterhaftigkeit der Sprache so zentral („phraseological tendency“ Sinclair 2004, 29), dass er nicht einzelne Wörter, sondern hauptsächlich größere Einheiten, die aus mehreren Wörtern bestehen, als Bedeutungsträger sieht. Damit wird auch Idiomatizität zu einem zentralen Phänomen (Sinclair 2004, 29f.) und Bedeutungen, die sich mit nur einem Wort verbalisieren lassen, die seltenen Ausnahmen der Sprache. Das bedeutet aber, dass jede frequente Wortkombination auch eine Kollokation ist.

Die terminologischen Varianten können immerhin in einem Punkt zusammengefasst werden: Von besonderem Interesse ist das Moment der Gebräuchlichkeit, der Typik (Feilke 1996, 72ff.) von sprachlichen Mustern. Und diese zeigt sich in der rekurrenten Struktur sprachlicher Einheiten. Dieses Phänomen kann mit dem Terminus der ‚Kookkur-

Determinationsverhältnis eines Kollokators (z. B. *schütteres*) zu einer Basis (*Haar*) postuliert wird.

<sup>8</sup> Im Original: ‚limited compositionality‘.

<sup>9</sup> In der Phraseologie würde dieses Phänomen als ‚Idiomatizität‘ bezeichnet, vgl. z. B. Burger (1998, 31f.).

renz‘ gefasst werden. Je nach terminologischer Verortung würden diese Phänomene (auch) als Kollokationen gefasst. Diese Definition deckt sich mit jener von Firth insofern, als die Gebräuchlichkeit einer Kookkurrenz als Maßstab dient, die die Wortverbindung zu einer Kollokation macht. Hierbei ist ebenso klar, dass dieser Maßstab von den Forschungsinteressen geleitet wird. Erstens kann er domänenspezifisch definiert werden: Kollokationen, die sich in einem allgemeinsprachlichen oder aber in einem thematisch oder textsortenspezifisch eingeschränkten Korpus als usuell zeigen. Zweitens ist die Definition von ‚gebräuchlich‘ oder ‚usuell‘ entscheidend für die Art des Maßstabs: Zeigt sich Gebräuchlichkeit durch hohe Frequenzen? Durch überzufälliges Aufeinandertreffen? Oder spielen Frequenzen eine untergeordnete Rolle zu Gunsten ganz anderer Faktoren?

Gebräuchlichkeit kann also unterschiedlich definiert werden. In der vorliegenden Untersuchung jedoch verlasse ich mich grundsätzlich auf Frequenz (und damit auf statistische Auffälligkeit), da diese den Überlegungen zur Musterhaftigkeit (vgl. Kapitel 2.2) am besten entsprechen.

Statistische Auffälligkeit kann unterschiedlich definiert werden. Grundsätzlich können aber zwei Gruppen von Definitionen ausgemacht werden:

1. In der einfachsten Definition werden alle möglichen Wortkombinationen (‚Kookkurrenzen‘) in einem Korpus gezählt. Je frequenter eine Wortkombination ist, desto eher wird die Wortkombination als musterhaft bezeichnet.
2. Mit statistischen Signifikanz-Tests werden die reinen Auftretensfrequenzen der Kollokationen in Beziehung zu den Eigenfrequenzen der Wörter, aus denen die Kollokation bestehen, gesetzt. Dabei gilt die Nullhypothese  $H_0$ : Dass zwei Wörter aufeinandertreffen ist zufällig. Wenn der statistische Test die Nullhypothese mit einer bestimmten Wahrscheinlichkeit verwirft, kann die Alternativhypothese  $H_1$  angenommen werden, die ein überzufälliges Aufeinandertreffen der Wörter postuliert.

Diese grundsätzliche Unterscheidung kann weiter verfeinert werden, wie Tabelle 6.1 auf der nächsten Seite zeigt. So können die absoluten

Signifikanz		Frequenzen	
		<i>absolut</i>	<i>relativ zur Korpusgröße</i>
Signifikanz	<i>nein</i>	Variante A: Frequenzen von Wortkombinationen in einem Korpus	Variante B: Frequenzen der Wortkombinationen aus A in Bezug zur Korpusgröße
	<i>ja</i>	Variante C: nicht sinnvoll	Variante D: Einbezug der Eigenfrequenzen der einzelnen Wörter der Wortgruppe; Berechnung der Überzufälligkeit

**Tabelle 6.1:** Grundsätzliche Möglichkeiten der Berechnung von Kollokationen.

Frequenzen der Kollokationen mit der Größe des Korpus in Beziehung gesetzt werden, um Vergleiche zwischen Korpora anstellen zu können.

Weiter können für alle Arten der Berechnung eine Reihe von Parametern festgesetzt werden:

1. **Länge der Wortgruppe:** Aus wie vielen Wörtern besteht die Wortgruppe? Mit Kollokationen sind oft Zweiwort-Kombinationen gemeint. Die Wortgruppe kann aber natürlich aus mehr als zwei Wörtern bestehen.
2. **Spannweite der Wortgruppe:** Entweder werden nur Wortkombinationen berücksichtigt, deren Token direkt aufeinander folgen (‘kontinuierliche Wortgruppen’). Oder aber es werden eine bestimmte Anzahl Token dazwischen erlaubt (‘diskontinuierliche Wortgruppen’, statt nur  $Wort_1 \rightarrow Wort_2$  auch  $Wort_1 \rightarrow [beliebige\ Wörter] \rightarrow Wort_2$ ).
3. **Reihenfolge der Wörter:** Entweder werden die Varianten  $Wort_1 \rightarrow Wort_2$  und  $Wort_2 \rightarrow Wort_1$  als zwei unterschiedliche Wortgruppen oder aber als Varianten derselben Wortgruppe gerechnet.
4. **Satzgrenzen:** Als Einschränkung kann definiert werden, dass eine Wortgruppe eine Satzgrenze nicht überschreiten darf.
5. **Suchbegriff:** Entweder werden bloß die Kollokationen zu einem bereits festgelegten Suchbegriff berechnet, oder aber es

ist das Ziel, alle Kollokationen in einem Korpus berechnen zu lassen.

Für bestimmte Definitionen mit entsprechenden Parametern haben sich je unterschiedliche Bezeichnungen ergeben. Dabei kommt es aber zu terminologischen Unschärfen, wie eingangs die schwierige Differenzierung zwischen Kookkurrenzen und Kollokationen bereits gezeigt hat. Ein weiterer verbreiteter Term ist der des ‚n-Gramms‘.<sup>10</sup> Das *n* steht für eine beliebige Zahl  $> 0$ ; die Bezeichnung leitet sich von den Namen für Ein-, Zwei- oder Dreiwortausdrücke, ‚Unigramme‘, ‚Bigramme‘, ‚Trigramme‘, ab. Normalerweise werden *n*-Gramme nur als eine Reihe von direkt aufeinander folgenden Wörtern verstanden (Manning/Schütze 2002, 192ff.). Neben dieser engeren Bedeutung kann aber auch von ‚diskontinuierlichen *n*-Grammen‘ gesprochen werden, wenn zwischen den Wörtern des *n*-Gramms auch Lücken erlaubt wären. Noch weiter gefasst wäre die Definition, wenn auch die Reihenfolge der Wörter frei sein könnte. Wortgruppen dieser Art bezeichnen Cheng u. a. (2006) mit dem Terminus ‚Concgram‘.

Eine weitere spezielle Form von Kollokationen sind ‚Collostructions‘, vorgeschlagen von Stefanowitsch/Gries (2003): Hier werden Kombinationen von Lexemen und morphosyntaktischen Konstruktionen untersucht. So untersuchen die Autoren beispielsweise die Bindungsstärke zwischen der grammatikalischen Konstruktion [N *waiting to happen*] mit verschiedenen ‚Collexemes‘ wie *accident*, *disaster* etc. und können so eine Bindung als ebenenübergreifend beschreiben.

## 6.2 Syntagmatische Muster

Am Institut für Deutsche Sprache (IDS) in Mannheim wurde eine Methode zur Berechnung von Kollokationen<sup>11</sup> und syntagmatischen Mustern entwickelt (Belica 2001–2006, Keibel/Belica 2007). Auf Grundlage des DeReKo IDS (o. J.)-Korpus erfolgt die Berechnung in mehreren Schritten (Institut für Deutsche Sprache 2004): Zunächst werden zu

<sup>10</sup> Ich deutsche mit dieser Schreibweise die englische Bezeichnung ‚*n*-grams‘ ein.

<sup>11</sup> Im DeReKo IDS (o. J.)-Korpus ist meist die Rede von ‚Kookkurrenzen‘, was auch als ‚Kollokation‘ bezeichnet werden könnte (vgl. Kapitel 6.1).

einer bestimmten Bezugseinheit, meist einem Wort, statistisch signifikante Kollokationen in einem definierten Kontext (z. B. fünf Wörter links und rechts der Bezugseinheit) ermittelt. Dann werden wiederum zu diesen ermittelten primären Kollokationspartnern der Bezugseinheit statistisch signifikante sekundäre Kookkurrenzen berechnet. Dieses Verfahren wird wiederholt, bis keine signifikanten Kontexte mehr gefunden werden. Schließlich werden die ermittelten Kollokationen zu einem syntagmatischen Muster vervollständigt. Dabei werden die typische Reihenfolge der Kollokationspartner zur Bezugseinheit, sowie die typischen Füllungen zwischen den Kollokationen ausgewertet. Damit ist für eine Bezugseinheit ein Kookkurrenzprofil erstellt worden: Ein Kookkurrenzprofil ist die Gesamtheit der „statistisch ermittelten Kookkurrenzcluster [...] für eine Bezugseinheit (z. B. für ein Bezugswort)“ (Belica/Steyer 2006, 10).<sup>12</sup>

Tabelle 6.2 auf der nächsten Seite zeigt Ausschnitte aus dem Kookkurrenzprofil des Lexems *Zahl*. Neben den Kollokationspartnern zu *Zahl* (vierte Spalte) sind der maximale linke und rechte Kontext in Anzahl Wörtern (erste und zweite Spalte), die Stärke der Beziehung der Kollokation als ‚log-likelihood-ratio‘ (LLR, dritte Spalte), die absoluten Frequenzen (fünfte Spalte), das syntagmatische Muster (sechste und siebte Spalte) mit einer Prozentangabe, in wie vielen Fällen dieses syntagmatische Muster gilt, aufgeführt.

Der erste Block mit den Zeilen 1–7 zeigt Varianten, bei denen *Arbeitslosen* stärkster Kollokationspartner zu *Zahl* ist. Weiter werden auch Kollokationspartner zur Kollokation *Zahl Arbeitslosen* angezeigt: Als komplexester Cluster ergibt sich *Zahl* in Kombination mit *Deutschland* und *registrierten* in der ersten Zeile mit allerdings bloß einem Beleg. Der einfachste Cluster (Zeile 7) kombiniert *Zahl* und *Arbeitslosen* ohne weitere statistisch signifikanten Kookkurrenzen mit 1900 Belegen.

In der vierten Spalte ist ersichtlich, wie diese Wörter normalerweise syntaktisch verankert sind: Beim letzten Cluster in Zeile 7 entsprechen 97 % der Fälle der Verwendung dem Muster *die Zahl der Arbeitslosen* mit Varianten in der Groß-/Kleinschreibung des bestimmten Artikels *die*.

<sup>12</sup> Sinclair (1996) etablierte die Bezeichnung „collocational profile“ (wieder abgedruckt in Sinclair 2004, zit. in Tognini-Bonelli 2001, 19).

Kontext		LLR	Kollokationspartner	Frequenz	syntagmatisches Muster		
links	rechts				Anteil	Beispiel	
1	2	2	18 553	Arbeitslosen registrierten Deuts	1	100 %	Zahl ... registrierten Arbeitslosen ... Deutschland
2	2	2	18 553	Arbeitslosen registrierten Milli	3	66 %	Millionen die Zahl der registrierten Arbeitslosen
3	2	2	18 553	Arbeitslosen registrierten	48	97 %	die! Die Zahl der registrierten [...] Arbeitslosen ... um
4	2	2	18 553	Arbeitslosen Deutschland Million	1	100 %	Millionen ... Zahl ... Arbeitslosen ... Deutschland
5	2	2	18 553	Arbeitslosen Deutschland	100	89 %	die! Die Zahl der Arbeitslosen in Deutschland ist
6	2	2	18 553	Arbeitslosen Millionen	85	41 %	die Zahl der Arbeitslosen auf vier Millionen
7	2	2	18 553	Arbeitslosen	1900	97 %	die! Die Zahl der Arbeitslosen
8	-1	-1	16 437	schwarze schreiben	545	88 %	wieder schwarze [...] Zahlen [...] schreiben
9	-1	-1	16 437	schwarze geschrieben	206	99 %	wieder schwarze [...] Zahlen [...] geschrieben werden
10	-1	-1	16 437	schwarze schreibt	207	62 %	schreibt [wieder] schwarze Zahlen
11	-1	-1	16 437	schwarze	1622	98 %	wieder schwarze [...] Zahlen schreiben
[Zeilen ausgelassen]							
19	-1	-1	13 234	roten	1398	97 %	in die! den roten [...] Zahlen
20	-3	5	13 059	stieg Todesopfer	57	66 %	Damit stieg die Zahl der Todesopfer auf ...
21	-3	5	13 059	stieg Übernachtungen	29	55 %	Die Zahl der Übernachtungen [...] stieg um ...
22	-3	5	13 059	stieg Fluggäste	18	55 %	stieg die Zahl der Fluggäste um ...
23	-3	5	13 059	stieg	2400	60 %	stieg [die] Zahl der ...
24	2	3	10 522	Beschäftigten sank	81	54 %	Die Zahl der Beschäftigten [...] sank um
25	2	3	10 522	Beschäftigten ging unselbständig	1	100 %	Zahl ... unselbständig Beschäftigten ging
26	2	3	10 522	Beschäftigten ging	57	50 %	Die Zahl der Beschäftigten [...] ging um
27	2	3	10 522	Beschäftigten unselbständig	29	100 %	Die! Die Zahl der unselbständig Beschäftigten um ...
28	2	3	10 522	Beschäftigten	1453	97 %	die! Die Zahl der Beschäftigten
[Zeilen ausgelassen]							
47	-4	5	6516	steigt stetig	25	36 %	Die Zahl [der ...] steigt [...] stetig
48	-4	5	6516	steigt kontinuierlich	22	54 %	Die! Die Zahl [der ...] steigt [...] kontinuierlich an
49	-4	5	6516	steigt ständig	42	52 %	Die! Die Zahl [der ...] steigt [...] ständig
50	-4	5	6516	steigt	1407	59 %	die! Die Zahl [der ...] steigt

Tabelle 6.2: Ausschnitt aus der Kookkurrenzdatenbank CCDB (Belica 2001–2006) zum Analysewort *Zahl*.

In den anderen syntagmatischen Mustern sind auch Auslassungspunkte als Füllungen ersichtlich: Sie stehen für eine zu heterogene Gestalt der Füllungen, als dass sie verallgemeinert werden könnten. Die Auslassungspunkte zeigen, dass die Wörter nicht unmittelbar aufeinander folgen, sondern immer („...“) oder manchmal („[...]“) eine syntagmatische Lücke aufweisen.

Wie erwähnt wird die Signifikanz der Kollokationen als log-likelihood-ratio ausgedrückt. Dieser Wert stellt ein Maß dar, mit dem die Abweichung der zufälligen Verteilung (das Wort befindet sich zufälligerweise im Kontext des Bezugsworts) von der beobachteten Verteilung ausgedrückt wird (Institut für Deutsche Sprache 2004). In anderen Arbeiten werden auch andere statistische Maße verwendet<sup>13</sup> oder die Kollokationen werden im einfachsten Fall nach Frequenzen geordnet.

Die syntagmatischen Muster stehen für eine Erweiterung des Kollokationen-Begriffs. Steyer (2004a, 34) verwendet deshalb folgerichtig auch eher die Bezeichnung „Mehrworteinheit“ oder „Mehrwortverbindung“, um nicht den Eindruck zu erwecken, es handle sich um einfache Bigramme.<sup>14</sup>

### 6.3 Zwischenfazit: Terminologische Festlegung

Die bisherigen Ausführungen haben gezeigt, dass eine große Vielfalt an unterschiedlichen Bezeichnungen für Wortgruppen im weitesten Sinn besteht.

Das terminologische Problem mit den Bezeichnungen „Kollokationen“ und „Kookkurrenzen“ löse ich in der vorliegenden Arbeit wie

<sup>13</sup> In Kapitel 7 werden eine Reihe von statistischen Tests dargestellt. Einen Überblick gebräuchlicher Maße geben auch Manning/Schütze (2002, 162ff.). Dort werden erwähnt: t-Test, Pearsons  $\chi^2$ -Test, Likelihood ratios, Relative frequency ratios, Mutual Information, z score, Exakter Test von Fisher.

<sup>14</sup> Die Analyse der syntagmatischen Muster hinsichtlich bestimmter Fragestellungen ist alles andere als trivial: Um in der Menge der vorhandenen Mehrworteinheiten Musterhaftes entdecken zu können, entwickelte Perkuhn (2007) ein „Vorgehensmodell“ zur systematischen Auswertung automatisch ermittelter Sprachmuster. Grob zusammengefasst werden damit (mit Hilfe eines entsprechenden Tools) die Kollokatoren zu einem Schlüsselwort dergestalt visualisiert, dass gleichzeitig Detailuntersuchungen und Überblicke möglich sind. Anschließend gibt es für die Interpretation Möglichkeiten der Annotation und Gruppierung der Kollokatoren.



folgt: Unter ‚Kollokationen‘ verstehe ich statistisch auffällige Kookkurrenzen. Das Plus von Kollokationen gegenüber Kookkurrenzen liegt also nur im statistischen Maß der überzufälligen Kombination. Kookkurrenzen beschreiben demnach in einem Syntagma aufeinander treffende Wörter, von denen nicht weiter bekannt ist, ob ihre Verbindung statistisch signifikant oder überhaupt besonders frequent ist.

Traditionellerweise bestehen Kollokationen aus jeweils zwei Wörtern. Die Bezeichnungen ‚n-Gramme‘, aber auch ‚syntagmatische Muster‘ oder ‚Mehrworteinheiten‘ erweitern solche Bigramme um beliebig viele Wörter. Auch im englischsprachigen Raum sind die Termini ‚Multi-Word Unit‘ oder ‚Multi-Word Expression‘ gebräuchlich, um einen flexibleren Kollokationen-Begriff zur Verfügung zu haben.<sup>15</sup> Um generell Wortverbindungen zu bezeichnen, die aus zwei oder mehr Wörtern bestehen, verwende ich deshalb auch den Terminus ‚Mehrworteinheit‘. Damit werden Wortkombinationen bezeichnet, die statistisch auffällig sind und aus beliebig vielen Wörtern in sowohl kontinuierlicher als auch diskontinuierlicher Reihe bestehen können.

Wenn die berechneten Mehrworteinheiten einem interpretatorischen Selektionsprozess unterzogen oder zu abstrakteren Einheiten zusammengefasst werden, spreche ich von ‚Mustern‘. Dieser Selektions- und Interpretationsprozess erfolgt genau dann, wenn die corpus-driven- in eine corpus-based-Perspektive umschlägt, in dem eine Mehrworteinheit hinsichtlich einer Fragestellung in den Fokus rückt. Oder kurz: Was die Maschine (unter meiner Definition der Algorithmen, Parameter und Daten) hervorbringt, sind (statistisch signifikante) Mehrworteinheiten. Die Menge der Mehrworteinheiten, die ich hinsichtlich des Forschungsinteresses eingrenze, systematisiere und beschreibe, sind Muster.

Somit wird auch klar, dass Konzepte wie Kollokationen, n-Gramme und Concgramme Varianten von Mehrworteinheiten sind, bei denen gewisse Parameter bereits festgelegte Werte aufweisen. Collostructions sind ebenfalls Varianten von Mehrworteinheiten, allerdings mit einer anders gelagerten Konzeption.

<sup>15</sup> Vgl. z. B. Oakes (1998, 184), Tognini-Bonelli (2001, 19), Rayson (2003, 35), Sinclair (2004, 31), Halliday u. a. (2004, 121).

## 6.4 Kollokationen, corpus-driven

Das vorherige Kapitel hat gezeigt, wie aus der Berechnung von Kollokationen zu einem Bezugswort syntagmatische Muster entstehen, die aus mehr als nur zwei Wörtern (Suchwort und Kollokat) bestehen können. Konsequenterweise muss dieser Vorgang für jedes Wort im Korpus ausgelöst werden.<sup>16</sup> Es gibt Verfahren, die das auf einfache Weise ermöglichen (vgl. z. B. Banerjee/Pedersen 2003, Cheng u. a. 2006).<sup>17</sup> In einem simplen Auszählverfahren werden die Frequenzen aller möglichen (nach Wunsch auch diskontinuierlicher) Wortgruppen berechnet. Zusätzlich können statistische Maße angewandt werden, um den Grad der Überzufälligkeit der Wortkombination auszudrücken.<sup>18</sup> Im Anhang in Kapitel A finden sich eine Reihe von so entstandenen Listen von häufigen Mehrworteinheiten, die die Analyse des NZZ-Korpus ergibt.<sup>19</sup>

Der Vorteil dieser Methode zeigt sich darin, dass sie weitgehend dem corpus-driven-Paradigma entspricht: Es ist der Versuch, Strukturen in einem Korpus sichtbar zu machen, ohne die Suche schon zu Beginn mit der Definition bestimmter Suchbegriffe einzuschränken. Mit diesem Versuch ist die Hoffnung verbunden, auf Strukturen zu stoßen, an die man zu Untersuchungsbeginn gar nicht gedacht hatte.<sup>20</sup>

<sup>16</sup> Mit der Kookkurrenzdatenbank CCDB (Belica 2001–2006) steht das Resultat einer solchen corpus-driven Berechnung auf Basis des DeReKo IDS (o. J.)-Korpus zur Verfügung. Allerdings wurden die Kookkurrenzprofile ‚bloß‘ von etwa 150 000 Lemmata berechnet. Und es ist nicht möglich, für ein neu definiertes Teilkorpus des IDS-Korpus oder für ein eigenes Korpus das Verfahren anzuwenden. Natürlich ist es möglich, über das Korpusrecherche- und -analysesystem COSMAS II (2007) zu einem beliebigen Suchausdruck ein Kookkurrenzprofil zu erstellen. Dies für jedes Wort des Korpus zu machen, ist jedoch praktisch unmöglich.

<sup>17</sup> Eine Zusammenstellung von Software, die das leistet, findet sich in Kapitel 9.3.

<sup>18</sup> Pearce (2002) bietet eine Gegenüberstellung unterschiedlicher statistischer Maße, die an einem „Gold-Standard“ von Kollokationen gemessen werden. Dieser Gold-Standard basiert auf Bigrammen aus dem New Oxford Dictionary of English. Der Autor weist allerdings selber auf das Problem der fehlenden eindeutigen Definition von ‚Kollokation‘ hin, so dass Gold-Standards je nach Untersuchungsinteresse unterschiedlich definiert werden müssen. Für meine Zwecke wäre der vom Autor gewählte Gold-Standard unpassend, da ich mich nicht nur für Kollokationen interessieren, die in einem Wörterbuch stehen. Eine weitere Gegenüberstellung von verschiedenen Signifikanzmaßen zur Berechnung von Kollokationen bietet Evert (2005).

<sup>19</sup> Die Listen im Anhang geben nur eine Auswahl von Mehrworteinheiten wieder. Die kompletten Listen sind online einsehbar: [www.bubenhofer.com/korpusanalyse/](http://www.bubenhofer.com/korpusanalyse/).

<sup>20</sup> Diese Hoffnung teilen z. B. auch Cheng u. a. (2006, 415), die mit der Software ‚Concgram‘ die flexibelste Form der Wortgruppenberechnung anbieten: „The fully automa-

Der Nachteil kann in technischen Beschränkungen liegen: Der Preis für flexible Definitionen von Kollokationen (Diskontinuität, freie Reihenfolge, Berücksichtigung des gesamten Korpus) ist hoch, was die nötige Rechenleistung betrifft. Und die entstehenden Datenmengen sind sehr umfangreich, so dass für die weitere Verarbeitung Verfahren der Filterung oder Fokussierung nötig sind. Auf solche Verfahren werde ich in Kapitel 8.1 zurück kommen.

### 6.5 Lemmata und Wortarten: Gewinn oder Verlust an Information?

Bis jetzt ging ich bei den beschriebenen Methoden der Berechnung von Kollokationen von einem nicht-annotierten Korpus aus. Das bedeutet, dass das Korpus nicht mit linguistischen Informationen angereichert wurde. Eine Annotation enthält z. B. das Lemma/den Type des Token oder Informationen über die Wortart. Die Computerlinguistik stellt Methoden zur Verfügung, um Korpora sehr rasch automatisch zu annotieren (sog. ‚Tagging‘ und ‚Parsing‘).<sup>21</sup>

Bei einem nicht-annotierten Korpus würden die Wortgruppen *im letzten Jahr* und *in den letzten Jahren* als je eigene Wortgruppen gezählt. Bei einem annotierten Korpus könnten statt der Token nur die Lemmata für die Zählung berücksichtigt werden. Dann würden die beiden Wortgruppen zusammengefasst zu *IN DER LETZTES JAHR*; Numerus und Kasus würden ignoriert.

Auf den ersten Blick scheint das Hinzufügen einer Information, hier also die Lemmatisierung, von Gewinn zu sein. Denn damit werden Wortgruppen mit den gleichen Lemmata auch als zusammengehörend erfasst und dadurch vielleicht erst zur statistisch signifikanten Wortgruppe. So könnte die Differenzierung von *einen großen Verlust*

---

ted capability of the search engine, i.e. the absence of any form of prior intervention by the user, makes it a truly ‚corpus-driven‘ methodology [...], and so further increases the likelihood that the conogram searches will enable the researcher to discover not only a more extensive description of patterns of collocation and their meanings, but also, and more importantly, new patterns of language use“.

<sup>21</sup> Vgl. für weiterführende Informationen Carstensen u. a. (2001, 373ff.). Ein auch für Zeitungstexte nützlicher Tagger ist der ‚TreeTagger‘ (Schmid 1994, 1995), mit dem ich experimentierte, um das in der vorliegenden Arbeit verwendete NZZ-Korpus zu annotieren.

FACING		FACED	
Bham Corpus	Economist & WSJ	Bham Corpus	Economist & WSJ
forwards	crunch	squarely	grim
palms	toughest	spun	dilemma
stood	challenges	dilemma	obstacles
sat	reformers	alternatives	challenges
problems	dilemma	prospect	prospect
windows	tasks	task	pressures
chair	troubles	choice	competition
sitting	choices	problem	difficulties
direction	difficulties	problems	shame
river	challenge	enormous	criticism
each	charges	challenge	choice
problem	task	practical	challenge
door	biggest	possibility	threat
table	prospect	situation	task
across	tough	threat	firms
feet	makers	with	tough
serious	immediate	crisis	defeat
wall	problems		
front	competition		

**Tabelle 6.3:** Kollokationen von *facing* und *faced* im ‚Birmingham Corpus‘ und im ‚Economist and Wall Street Journal Corpus‘ (Tognini-Bonelli 2001, 94).

und *ein großer Verlust* unerwünscht und stattdessen die lemmatisierte Form EIN GROSS VERLUST für die Analyse sinnvoll sein.

Es gibt aber gute Gründe, gegenteilig zu argumentieren: Die Lemmatisierung, und auch die Annotation mit Wortarten, ist ein Informationsverlust: Tognini-Bonelli (2001, 92) zeigt an einem Beispiel, dass die Vorstellung, ein Lemma bilde ein semantisches Paradigma, unsinnig ist. Am Beispiel von *facing* vs. *faced* wird das deutlich. Eine Korpusanalyse in einem allgemeinsprachlichen Korpus und einem Korpus aus Zeitungsartikeln der Wirtschaftspresse zeigt differenzierte Gebrauchsweisen (vgl. Tabelle 6.3).

Die Kollokationen von *facing* weisen im Birmingham Corpus in den meisten Fällen auf eine konkrete Verwendung hin, mit der die Position eines Objekts beschrieben wird. Nur wenige Kollokatoren weisen auf eine abstrakte Verwendung hin. Anders im Economist/WSJ-Corpus: Hier sind es alles abstrakte Verwendungsweisen von *facing*. Im Fall von *faced* sieht das Bild wiederum anders aus. In beiden Korpora finden sich nur abstrakte Verwendungsweisen.

Das Beispiel zeigt, wie mit der Lemmatisierung wichtige Gebrauchsnancen verloren gehen könnten:

One glance at the collocational profiles [...] dispels any possible illusion that inflected forms are grammatical variations of a certain base form, but broadly share the same meaning of the base form and have a similar behaviour. Elements from the context in which these two forms operate are found in their verbal co-text and so we find traces of the participants in the verbal event, the verbal action, and the non-verbal events connected with it. (Tognini-Bonelli 2001, 94)

Doch damit nicht genug: Tognini-Bonelli (2001, 99) geht mit Sinclair (1991, 7) überein und betont, dass die Unterscheidung von Form und Bedeutung nur als methodisches Kriterium einer bestimmten Perspektive auf Sprache dient, die primär *Langue* statt *Parole* im Blick hat. Dabei wird die Form jedoch zu Unrecht vernachlässigt:

Soon it was realised that form could actually be a determiner of meaning, and a causal connection was postulated, inviting arguments from form to meaning. Then a conceptual adjustment was made, with the realization that the choice of a meaning, anywhere in a text, must have a profound effect on the surrounding choices. It would be futile to imagine otherwise. There is ultimately no distinction between form and meaning. (Sinclair 1991, 7)

Somit wird klar, dass die Annotation eines Korpus mit Lemma- und Wortarteninformationen den Informationsgehalt reduziert (Sinclair 1992, 385). Werden bei der Analyse dann nur die Lemmata und/oder Wortarten berücksichtigt, macht sich dies negativ bemerkbar, da die Differenzierungsmerkmale, die sich in der morphosyntaktischen Ausprägung der Wörter zeigen, verloren gehen.

Natürlich kann hier eingewandt werden, dass die ursprüngliche Information der Rohdaten nicht gelöscht wird und weiterhin damit gearbeitet werden kann. Doch die Gefahr dabei ist groß, die Rohdaten zu ignorieren und nur noch die Strukturen der (gemäß einer bestimmten Theorie) nachbearbeiteten Daten zu fokussieren:

It could be argued that in a tagged text no information is lost because the words of the text are still there and available, but the problem is that they are bypassed in the normal use of a tagged text. The actual loss of information takes place when, once the annotation of the corpus is completed and the tagsets are attached to the data, the linguist processes the tags rather than the raw data. By doing this the linguist will easily lose sight of the contextual features associated with a certain item and will accept single, uni-functional items – tags – as the primary data. What is lost, therefore, is the ability to analyse the inherent variability of language which is realised in the very tight interconnection between lexical and grammatical patterns. (Tognini-Bonelli 2001, 73)

Nebenbei wird sichtbar, wie sich dieser Informationsverlust beim weiteren Parsing des Textes bemerkbar macht. So entstehen durch die Rückführung der Token auf Lemmata neue Ambiguitäten bei homographen Ausdrücken: In einem englischen Text kann ein lemmatisiertes ‚lie‘ z. B. nicht sicher als Infinitiv von *liegen* (*lay*, *lain*) oder *lügen* (*lied*, *lied*) erkannt werden. Im flektierten Kontext ist dies jedoch eher möglich, wie Tognini-Bonelli (2001, 74) zu recht feststellt. Immerhin kann da entgegnet werden, dass Tagger und Parser schon lange Kontexte im Rahmen von Wahrscheinlichkeitsmodellen, und somit konkrete Sprachgebräuche, in die Analyse miteinbeziehen.<sup>22</sup>

Auch Volk (2001, 157f.) bestätigt das Problem, dass mit der Lemmatisierung Bedeutungsnuancen und damit wichtige Informationen verloren gehen können. Er arbeitet mit einem Wahrscheinlichkeitsmodell, um beim maschinellen Parsen von Sätzen entscheiden zu können, ob eine Präpositionalphrase (*mit dem Fernglas*) an die Verbalphrase (*[sie] sieht*) oder die Nominalphrase (*den Mann*) gebunden werden muss. Das Modell berechnet die Wahrscheinlichkeiten mit der in Korpora Nomen bzw. Verben und Präpositionen kookkurrieren. Die gefundenen Häufigkeiten entscheiden über die Anbindung der Präpositionalphrase. Es zeigt sich, dass bei der Verwendung der Lemmata anstelle der Wortformen das Wahrscheinlichkeitsmodell weniger genau ist, mitunter deshalb, weil die unterschiedlichen Wortformen eines Paradigmas mit unterschiedlichen Präpositionen oder überhaupt nicht mit Präpositionen kookkurrieren. Volk (2001, 157f.) zeigt anhand von Analysen im WWW, dass z. B. *Verhandlungen* in vielen Fällen mit der Präposition *mit* kookkurriert, während der Singular, *Verhandlung*, kaum mit dieser Präposition verwendet wird. Das liegt daran, dass *Verhandlung* im Singular meistens in der Bedeutung ‚Gerichtsverhandlung‘ verwendet wird, während es meistens die Pluralform *Verhandlungen* ist, die in Ausdrücken wie *Verhandlungen mit jemanden führen* erscheint.

Ich werde für die Beispielanalysen der vorliegenden Arbeit also nicht mit einem mit Wortarten und Lemmata annotierten Korpus arbeiten, zeige aber im Schlusskapitel anhand eines anderen Korpus (‚Politikforum.de-Korpus‘) im Sinne eines Ausblicks, welche Analysemöglichkeiten sich bei einem entsprechend annotierten Korpus

<sup>22</sup> Vgl. hierzu z. B. die Funktionsweise des TreeTaggers (Schmid 1994).

ergeben (vgl. Kapitel 17.1). Allerdings zeigt sich auch bei diesem annotierten Korpus, dass die Berechnung der signifikanten Mehrworteinheiten mit Vorteil nicht auf den Lemmata, sondern auf den laufenden Wortformen basiert. Denn so können Ambiguitäten vermieden werden, die z. B. bei lemmatisierten Mehrworteinheiten wie *im mittlerer Weste/Westen* entstehen, weil die Wortform *Westen* sowohl die Pluralform des Kleidungsstück *Weste* als auch eine Himmelsrichtung sein kann. Der Tagger schlägt deshalb korrekterweise beide Lemmata vor, was bei der Analyse der Mehrworteinheiten jedoch zu Unsicherheiten führt. Die nicht-lemmatisierte Mehrworteinheit *im mittleren Westen* leuchtet dagegen unmittelbar ein.

Wenn sich diese Unsicherheiten bei der Analyse auch rasch beseitigen lassen, indem beide Lemma-Formen auf ihre Plausibilität hin überprüft werden, wirkt sich bei der Berechnung der Mehrworteinheiten auf Basis der Lemmata die bereits oben erwähnte Zusammenfassung von unterschiedlichen morphosyntaktischen Formen fatal aus: Listen von so berechneten Mehrworteinheiten unterscheiden sich stark von solchen, die auf Basis der laufenden Wortformen entstanden sind, wie Beispielberechnungen in einem annotierten Korpus zeigen.<sup>23</sup> Die oben referierten Überlegungen lassen vermuten, dass in der Liste der lemmatisierten Formen durch die Zusammenfassung der unterschiedlichen Wortformen wichtige Unterscheidungen verloren gegangen sind.

Die Überlegungen zum Informationsverlust und zur unerwünschten theoretischen Vorfixierung durch das Tagging von Korpora können auf das Problem der Modellierung von sprachlichen Mustern ganz allgemein übertragen werden. Es ist zumindest fragwürdig, wie stark ein modelliertes sprachliches Muster verallgemeinert werden soll (z. B. hinsichtlich der Flexionsformen oder durch Abstrahierungen der Füllungen von Formeln nach dem Muster *IN DER NACHT AUF [WOCHENTAG]*), ohne dass wichtige Informationen über die konkrete Verwendung in Texten verloren gehen. Ebenso kritisch könnte sich der Parameter ‚Reihenfolge der Wörter‘ bei der Berechnung von Kolloka-

<sup>23</sup> Es handelt sich um das oben erwähnte ‚Politikforum.de-Korpus‘, das in Kapitel 17.1 beschrieben wird. Die Testlisten sind online verfügbar: [www.bubenhofer.com/korpusanalyse/Politikforum.de/VergleichLemmaToken/](http://www.bubenhofer.com/korpusanalyse/Politikforum.de/VergleichLemmaToken/).

tionen erweisen. In den unterschiedlichen Abfolgen der zwar immer gleichen Wörter liegen wahrscheinlich auch Bedeutungsnuancen.<sup>24</sup>

Neben diesen kritischen Einwänden gegenüber der Lemmatisierung gibt es aber auch klare Vorteile. Denn die Annotation mit Wortarten würde es ermöglichen, ‚Collostructions‘ zu berechnen, also Mehrwort-einheiten, die sich aus bestimmten laufenden Wörtern, Lemmata und Wortartklassen zusammensetzen können (vgl. Stefanowitsch/Gries 2003 und Seite 118 in Kapitel 6.1 der vorliegenden Arbeit). Weiter ergeben sich Vorteile bei der corpus-based Recherche im Korpus, wenn direkt nach Lemmata oder Wortarten gesucht werden kann. Immerhin kann eine flexible Suchanfragesprache (etwa unter Verwendung von regulären Ausdrücken) dies in einem nicht-annotierten Korpus ebenfalls weitgehend ermöglichen (vgl. dazu auch Kapitel 9.2).

---

<sup>24</sup> Aspekte dieses Phänomens werden in der Grammatik mit den klassischen Termini der Herausstellungsstrukturen (also dem Mittel, syntaktische Elemente an, vor oder hinter die Satzgrenze zu verschieben) oder der Topikalisierung (der Vorfeldbesetzung) beschrieben. Diese bieten letztlich auch eine semantische Differenzierungsmöglichkeit.





## 7 Der statistische Zugang zu sprachlichen Daten

Die mathematische Teildisziplin der Statistik verfügt über einen umfangreichen Methodenapparat, um den Informationswert von Datenmengen zusammenzufassen und Hypothesen über Zusammenhänge innerhalb der Daten zu überprüfen. Es ist naheliegend, diese Methoden für die Korpuslinguistik fruchtbar zu machen. Dies beginnt bei deskriptiven statistischen Verfahren, wie dem Berechnen von Durchschnitten (‚arithmetisches Mittel‘ oder ‚Median‘) und der Darstellung von Häufigkeitsverteilungen, und reicht bis zu komplexeren Verfahren, wie den Signifikanztests, die die Wahrscheinlichkeit einer Abhängigkeit zwischen Variablen überprüfen.

Es geht in diesem Kapitel nicht darum, die Grundlagen der Statistik darzulegen.<sup>1</sup> Stattdessen stelle ich hier einige wenige Verfahren vor, die in der vorliegenden Arbeit für folgende Probleme angewandt werden:

1. Kollokationen berechnen: Die Wahrscheinlichkeit, dass zwei Wörter oft aufeinander treffen, hängt von den Frequenzen ab, mit denen die Wörter einzeln im Korpus verwendet werden. Statistische Tests können bei der Berechnung von Kollokationen aussagen, wie überraschend es ist, dass zwei Wörter zusammen auftreten.
2. Verteilungen prüfen: Es werden die Verteilungen bestimmter Wörter in zwei oder mehr Teilkorpora verglichen. Wenn beobachtet wird, dass ein bestimmtes Element in mehreren Teilkorpora im Vergleich unterschiedlich oft erscheint, kann getestet werden, ob diese Unterschiede genügend groß sind, um sie nicht als zufällige Schwankungen interpretieren zu müssen.
3. Korpora vergleichen: Aufgrund bestimmter Merkmale, z. B. der Frequenzen von definierten Elementen, können zwei Korpo-

---

<sup>1</sup> Dazu sei auf Toutenburg (2000a,b), sowie speziell für linguistische Fragestellungen auf Albert/Koster (2002), Oakes (1998), Rietveld/Hout (2005) und Woods u. a. (1986) verwiesen. Anleitungen für die korpuslinguistische Arbeit mit dem Statistikpaket ‚R‘ bietet Baayen (2008).

ra miteinander verglichen werden. Dabei stellt sich auch hier die Frage, welche Merkmale genügend verschieden sind, damit nicht von zufälligen Schwankungen gesprochen werden muss. Dieser Anwendungsbereich ist bloß eine Verallgemeinerung von Punkt 2.

Um die genannten Probleme lösen zu können, werden einerseits Verfahren der deskriptiven Statistik angewandt, die im folgenden Kapitel dargestellt werden. Andererseits müssen Hypothesen über die Signifikanz von Korrelationen zwischen Variablen überprüft werden. Dazu dienen statistische Tests, die in Kapitel 7.2 genannt werden.

### 7.1 Deskriptive Statistik

In der deskriptiven Statistik wird versucht, die Daten einer Stichprobe zusammenzufassen. Dies geschieht durch die Angabe von absoluten oder (im vorliegenden Fall z. B. zur Korpusgröße) relativen Häufigkeiten (Toutenburg 2000a, 21ff.), z. B. von Kollokationen, oder aber durch sog. Maßzahlen. Diese unterteilen sich in Lagemaße und Streuungsmaße (Toutenburg 2000a, 43ff.). Zu den Lagemaßen gehören das arithmetische Mittel, der Median und weitere Maße, die für den vorliegenden Anwendungsbereich nicht von Interesse sind.

Die Streuungsmaße erlauben Aussagen über die Art des Bereichs, in der sich die Häufigkeitsverteilung bewegt. Mit der Varianz  $s^2$  und der Standardabweichung  $s$  wird berechnet, wie stark die einzelnen Werte einer Messung vom Mittel abweichen. Dabei berechnet sich die Standardabweichung für die Messwerte  $x$  wie folgt:

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (7.1)$$

In der Korpuslinguistik kann die Berechnung der Standardabweichung u. a. für folgende Zwecke eingesetzt werden:

1. Distribution: Wenn eine Messreihe von Frequenzen eines Phänomens zu unterschiedlichen Zeitpunkten oder in mehreren

Teilkorpora besteht, wird mit  $s$  angegeben, wie stark die einzelnen Frequenzen vom Mittelwert abweichen. Bei einer Messreihe über verschiedene Jahre würde eine tiefe Standardabweichung bedeuten, dass es keine großen Schwankungen in der Verwendung des gemessenen Phänomens gibt.

2. Kollokationen: Bei der Berechnung von diskontinuierlichen Kollokationen kann z. B. nicht nur beschrieben werden, wie weit die Token im Mittel voneinander entfernt sind, sondern wie stark die einzelnen Distanzen vom Mittel abweichen (Manning/Schütze 2002, 158f.). Die Formel lautet dann leicht modifiziert:

$$s = \sqrt{\frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n - 1}} \quad (7.2)$$

$n$  ist die Frequenz der Kookkurrenz,  $d_i$  die Distanz zwischen den Token der Kookkurrenz  $i$  und  $\bar{d}$  das arithmetische Mittel aller Distanzen. Eine tiefe Standardabweichung bedeutet, dass die Token immer etwa in der gleichen Distanz im Korpus vorkommen.

Da die Standardabweichung von der Höhe des Mittelwerts abhängt, kann zusätzlich der Variationskoeffizient  $V$  berechnet werden. Dieses Maß, das die Standardabweichung  $s$  in Relation zum Mittelwert setzt ( $s/\bar{x}$ ), ermöglicht einen Vergleich der Variation von verschiedenen Datensätzen, deren Frequenzen sich auf unterschiedlichen Niveaus bewegen.<sup>2</sup>

<sup>2</sup> Bei einer angenommenen Datenreihe  $A = \{5, 4, 4, 3, 8, 5, 5\}$  ergibt sich eine Standardabweichung von  $s = 1,57$ , bei einer Datenreihe  $B = \{320, 323, 340, 315, 320, 340, 325\}$  ist  $s = 9,69$ . Die Abweichungen in der zweiten Datenreihe sind aber weniger überraschend als in der ersten Datenreihe. Deshalb setzt der Variationskoeffizient die Standardabweichung in Relation zum Mittelwert, was bei  $A$  einen Koeffizienten von  $V = 0,32$  und bei  $B$  von  $V = 0,03$  gibt.

## 7.2 Testen von Hypothesen

Letztlich gibt es für quantitative Analysen jeweils dasselbe statistische Kernproblem: Nämlich die Fragen, ob es eine Korrelation zwischen zwei Variablen gibt und ob diese Korrelation statistisch signifikant ist, wenn man die Größe der Stichprobe berücksichtigt.

Die Korrelation drückt aus, ob es einen Zusammenhang zwischen zwei Variablen gibt.<sup>3</sup> Ein solcher Zusammenhang wird allgemein-sprachlich etwa so ausgedrückt: Immer wenn X auftritt, tritt auch Y auf. Oder: Je größer X ist, desto größer ist auch Y.

Es gibt zahlreiche Tests, um Korrelationen und ihre Signifikanz zu prüfen. Welche Tests verwendet werden können, hängt von der Art der Daten ab, etwa der Messskala (Nominal-, Ordinal- oder metrische Skala)<sup>4</sup> oder des Verteilungstyps der Variablen (bekannte vs. unbekannte Verteilung).<sup>5</sup> Allerdings spielen im Kontext der Korpus-linguistik weitere Faktoren bei der Testauswahl eine Rolle.

### 7.2.1 Klumpen: Wörter sind nie zufällig verteilt

Mit Sprache als Datenmaterial ergeben sich spezifische Probleme bei der Anwendung der Tests, wie Kilgarriff (2005) darlegt: Die Verteilung von Wörtern im Korpus ist nicht zufällig. Als „clumpiness“ oder „burstiness“ (Church/Gale 1995; Kilgarriff 2001, 107), als ‚Klumpen‘ oder ‚Häufungen‘ also wird das Phänomen beschrieben:

Where a word occurs once in a text, you are substantially more likely to see it again than if it had not occurred once. Once a corpus is seen as having internal structure – that is, comprising distinct texts – the independence assumption is unsustainable. (Kilgarriff 2001, 107)

Damit wird deutlich, dass die Nullhypothese  $H_0$  bei genug großem Datenumfang immer verworfen wird, wenn ein Test die Zufälligkeit der Verteilung der Daten misst. Dies beweist Kilgarriff (2001) experimentell,<sup>6</sup> u. a. mit einer Wiederholung des Korpusvergleichs Amerikanischen und Britischen Englischs:

<sup>3</sup> Dabei wird zwischen ‚Korrelation‘ und ‚Assoziation‘ unterschieden: Während die Assoziation einen beliebigen Zusammenhang ausdrückt, steht Korrelation für eine lineare Beziehung (Toutenburg 2000a, 121). In der Forschungspraxis wird jedoch oft für Assoziationen auch der Begriff Korrelation verwendet.

<sup>4</sup> Vgl. Toutenburg (2000a, 6), Albert/Koster (2002, 74ff.).

<sup>5</sup> Vgl. Toutenburg (2000b, 109ff.), Oakes (1998, 10f.).

<sup>6</sup> Vgl. auch Kilgarriff (1996a,b), Kilgarriff/Rose (1998).

The results were as follows. For 3,418 of the words, the null hypothesis was defeated (at a 97,5 % significance level). In corpus statistics, this sort of result is not surprising. Few words comply with the null hypothesis, but then, as discussed above, the null hypothesis has little appeal: there is no intrinsic reason to expect any word to have exactly the same frequency of occurrence on both sides of the Atlantic. We are not in fact concerned with whether the null hypothesis holds: rather, we are interested in the words that are furthest from it. (Kilgarriff 2001, 114)

Die Lösung des Problems liegt also in den folgenden Möglichkeiten:

1. Verwendung eines statistischen Tests, der weniger stark auf Klumpen innerhalb einer Stichprobe reagiert. Kilgarriff (2001, 103) schlägt den Mann-Whitney-Rank-Test vor.
2. Ein statistischer Test, wie der  $\chi^2$ -Test, darf nicht dafür verwendet werden, über die Signifikanz einer Verteilung zu entscheiden, sondern bloß über den Grad der Unzufälligkeit der Verteilung.<sup>7</sup>

Es gibt aber einen weiteren Ausweg: Bis jetzt bin ich davon ausgegangen, dass im Korpus die Vorkommen von Ausdrücken gezählt werden. Die Einheit ‚Text‘ spielt dabei keine Rolle. Ob ein bestimmter Ausdruck zwar viele Male, aber immer im selben Text vorkommt, oder aber über mehrere Texte verteilt ist, wird nicht berücksichtigt. Wenn das Korpus jedoch in Texte gegliedert ist, wie das beim NZZ-Korpus der Fall ist, können statt Ausdrücke Artikel gezählt werden, in denen der Ausdruck mindestens einmal erscheint. Es werden dann Artikel pro Jahr, Monat oder Tag gezählt, in denen der Suchausdruck vorkommt. Dabei spielt keine Rolle, wie oft der Ausdruck im Artikel zu finden ist; es zählt nur die Entscheidung, ob in einem Artikel der Suchausdruck mindestens einmal vorkommt oder gar nicht. Damit wird verhindert, dass eine Häufung eines Ausdrucks in einem Artikel übermäßig stark gewichtet wird. Das Problem der Klumpenhaftigkeit entschärft sich damit etwas, wobei natürlich nach wie vor zu erwarten ist, dass aufgrund der strukturellen Eigenschaften des Mediums ‚Zeitung‘ bezüglich Thema oder Textsorte ähnliche Artikel direkt aufeinander folgen und daher auch der Sprachgebrauch klumpenhaft ist.

<sup>7</sup> Vgl. dazu den Ausschnitt aus dem obigen Zitat: „We are not in fact concerned with whether the null hypothesis holds: rather, we are interested in the words that are furthest from it“ (Kilgarriff 2001, 114).

	S	$\neg S$	Total
P	A	B	A + B
$\neg P$	C	D	C + D
Total	A + C	B + D	N

**Tabelle 7.1:** Kontingenztafel für zwei Variablen S und P ( $\neg$  ist der Negationsoperator).

Es ist ratsam, bei den Analysen jeweils die Berechnungen auf beiden Grundlagen vorzunehmen, also auf der Basis von Wörtern und auf der Basis von Artikeln. Bestehen hier große Unterschiede, müssen diese genauer analysiert werden.

In der Folge zeige ich, welche statistischen Testverfahren für die vorliegende Untersuchung verwendet werden können.

### 7.2.2 Kontingenztafel erstellen

Für die meisten Tests wird zunächst eine Kontingenztafel erstellt, die die beobachteten und erwarteten Frequenzen für alle Fallkategorien enthält (vgl. Tabelle 7.1). Um die Signifikanz einer Zwei-Wort-Kollokation zu testen (z. B. *runder Tisch*, werden die beiden Wörter den Variablen P und S zugewiesen. Feld A enthält die Anzahl der Fälle, bei denen Wort P auf Wort S trifft (*runder* + *Tisch*), Feld B die Fälle, in denen P auftritt, jedoch nicht gleichzeitig S (*runder* + X [aber nicht *Tisch*]), Feld C die Fälle, in denen zwar S nicht aber P auftritt (X [aber nicht *runder*] + *Tisch*) und Feld D, in denen weder P noch S auftreten. Entsprechend berechnen sich die jeweiligen Totalen, N ist die Zahl aller Fälle.

Die Kontingenztafel kann jedoch auch verwendet werden, um die Unterschiede der Frequenzen eines Wortes oder einer Wortgruppe in zwei Korpora auf ihre Signifikanz zu testen. Die Variable P steht dann für das Korpus: P wäre Korpus A,  $\neg P$  Korpus B. Die Variable S steht für das Phänomen, also das Wort bzw. die Wortgruppe: Die Felder A und B enthalten die Anzahl Fälle des Phänomens in Korpus A und B, die Felder C und D die Frequenzen aller anderen Fälle.

Die oben dargestellte Kontingenztafel ist eine Vierfeldertafel und damit der einfachste Typ von Kontingenztafeln. Die Tafel kann jedoch beliebig viele Felder aufweisen, so würde für den Signifikanztest der

Verteilung einer Kollokation in den Jahren von 1995 bis 2005 eine  $11 \times 2$ -Feldertafel benötigt.

### 7.2.3 Chi-Quadrat-Test

#### Definition

Der Chi-Quadrat-Test (im Folgenden: „ $\chi^2$ -Test“) prüft die beobachteten Frequenzen in einer Kontingenztafel mit den erwarteten Werten auf Unabhängigkeit.

$$\chi^2 = \sum \frac{(O - E)^2}{E} \quad (7.3)$$

Für jede Zelle der Kontingenztafel wird

$$\frac{(O - E)^2}{E} \quad (7.4)$$

berechnet, wobei O für die beobachtete, E für die erwartete Frequenz steht. Für jede Zelle berechnet sich E nach folgender Formel:

$$\frac{\text{Reihensumme} \times \text{Spaltensumme}}{\text{Total aller Frequenzen}} \quad (7.5)$$

Jetzt kann geprüft werden, mit welcher Wahrscheinlichkeit p die Nullhypothese  $H_0$ , die eine zufällige Verteilung behauptet, verworfen werden kann. Dafür müssen die Freiheitsgrade df bestimmt werden, die sich wie folgt berechnen lassen:  $(\text{Reihenzahl} - 1) \times (\text{Spaltenzahl} - 1)$ . Bei einer  $2 \times 2$ -Tabelle beträgt df also 1. Die Tabelle der kritischen Werte für  $\chi^2$  gibt dann Auskunft über die minimale Höhe, die  $\chi^2$  für ein bestimmtes Signifikanzniveau haben muss, um  $H_0$  verwerfen zu können.

Eine Besonderheit ergibt sich für Kontingenztafeln mit nur zwei Reihen und Spalten (also mit  $df = 1$ ) oder mit sehr kleinen Werten. Dort sollte  $\chi^2$  unter Benutzung der Yates-Korrektur verwendet werden:

$$\chi^2 = \sum \frac{(|O - E| - 0,5)^2}{E} \quad (7.6)$$



Für den Zähler wird also vom absoluten Wert<sup>8</sup> der Differenz von O und E 0,5 abgezogen.

$\chi^2$  gibt zwar Auskunft darüber, ob eine Beziehung zwischen Variablen existiert, jedoch nicht direkt, wie stark sie ist. Denn der Maximalwert von  $\chi^2$  hängt von N und der Größe der Kontingenztafel ab.

Um  $\chi^2$  direkt in einem standardisierten Maß auszudrücken, stehen verschiedene Möglichkeiten offen. Darunter fallen der Phi-Koeffizient

$$\phi = \sqrt{\frac{\chi^2}{N}} \quad (7.7)$$

und das Kontingenzmaß von Cramér, das den Phi-Koeffizienten um die Dimension der Kontingenztafel bereinigt:

$$\text{Cramér's } V = \sqrt{\frac{\chi^2}{N(k-1)}} \quad (7.8)$$

Dabei ist N die Gesamtzahl der beobachteten Werte und k die kleinere Zahl der Spalten oder Reihen der Kontingenztafel. Dabei gilt:  $0 \leq V \leq 1$ ; ein V nahe 0 zeigt eine schwache, nahe 1 eine starke Beziehung an.

### Diskussion

Für diesen Test gilt das oben in Kapitel 7.2.1 dargestellte Problem der Normalverteilung. Zwar handelt es sich um einen nicht-parametrischen Test, bei dem nicht verlangt wird, dass die Stichprobe normalverteilt ist; allerdings erwartet er es von der Grundgesamtheit. Um Vergleiche zwischen Korpora anzustellen, postuliert Kilgarriff (2001, 123) den Test nicht als Hypothesen-Test zu verwenden, da die Nullhypothese bei Sprachdaten in größerem Umfang beinahe immer verworfen wird (Kilgarriff 2001, 99ff.). Hingegen kann der Test verwendet werden, um im Vergleich zu anderen Kollokationen im selben Korpus oder zu selben Kollokationen in anderen Korpora eine Rangordnung zu erzeugen. Im Vergleich mit anderen Statistiken zeigt Kilgarriff (2001, 121ff.), dass der  $\chi^2$ -Test für den Vergleich von Korpora am Präzisesten ist. Allgemein wird der Test bei kleinen beobachteten Frequenzen (unter 5) unzuverlässig.

<sup>8</sup> Nach der Subtraktion wird das Vorzeichen positiv gesetzt, sollte es negativ sein.

#### 7.2.4 Log-Likelihood $G^2$

##### Definition

Auch der Log-Likelihood-Koeffizient berechnet aufgrund einer Vierfelder-Kontingenztafel das Verhältnis zwischen zwei Modellen:

$$\begin{aligned}
 G^2 = 2[ & A \log A + B \log B + C \log C + D \log D \\
 & - (A + B) \log(A + B) - (A + C) \log(A + C) \\
 & - (B + D) \log(B + D) - (C + D) \log(C + D) \\
 & + (A + B + C + D) \log(A + B + C + D)] \quad (7.9)
 \end{aligned}$$

Der Test folgt auch der  $\chi^2$ -Verteilung. Deshalb können die kritischen Werte von  $\chi^2$  für den Hypothesentest verwendet werden (Manning/Schütze 2002, 174).

##### Diskussion

Obwohl auch dieser Test von einer Normalverteilung ausgeht, schlägt ihn Dunning (1993) vor, um auch seltene Ereignisse als signifikant zu gewichten, was auf viele Kollokationen in einem Korpus zutrifft. Der Test ist damit im Vorteil gegenüber dem  $\chi^2$ -Test, der bei geringen Fallzahlen unzuverlässig wird. Bei hohen Fallzahlen sind jedoch die Ergebnisse des  $\chi^2$ - und des  $G^2$ -Tests gleich zuverlässig (Oakes 1998, 38). Manning/Schütze (2002, 172ff.) heben den Vorteil der einfachen Interpretation des Maßes hervor. Rayson/Garside (2000) verwenden den Test für Korpusvergleiche und auch Kilgarriff (2001, 105) befindet ihn zumindest für kleine und mittlere Frequenzen von Wörtern als gut, um Vergleiche zwischen Korpora zu ziehen.

#### 7.2.5 Weitere Tests

##### Mann-Whitney-Rank-Test

Der Mann-Whitney-Rank-Test, der auch unter der Bezeichnung Wilcoxon-Rank-Test (bzw. U-Test) bekannt ist, kann in der Korpuslinguistik zum Vergleich der Verteilung von sprachlichen Einheiten in zwei Korpora dienen. Im Unterschied zu Tests, die eine Zufallsverteilung erwarten, kann mit dem Mann-Whitney-Rank-Test das Problem

Frequenz:	3	3	12	13	15	18	24	27	33	88	Total
Korpus:	Y	Y	X	Y	X	X	X	Y	Y	X	
Rang (X)			3		5	6	7			10	31
Rang (Y)	1	2		4				8	9		24

**Tabelle 7.2:** Der Mann-Whitney-Rank-Test für zwei Korpora X und Y (Kilgarriff 2001, 103).

der klumpenhaften Verteilung von Wörtern in Sprache umgangen werden, wie Kilgarriff (2001, 103)<sup>9</sup> vorschlägt. Denn die Korpora werden in eine bestimmte Anzahl gleich großer Abschnitte unterteilt, wobei für jeden Teil die Frequenzen berechnet werden. Statt nun aber direkt die Differenzen der Frequenzen zwischen den Korpora zu vergleichen, werden die Frequenzen in eine Rangliste gebracht, worin auch die Herkunft jeder Frequenz vermerkt wird.

Die Tabelle 7.2 zeigt eine Rangliste von Frequenzen aus den Korpora X und Y. Die beiden Korpora wurden in je fünf Teile geteilt. Für jeden Teil wurde die Frequenz des Suchausdrucks berechnet. Die beiden letzten Zeilen der Tabelle erfassen die Ränge und ihre Zuordnung zu den beiden Korpora.

Der Test prüft dann die Nullhypothese (alle Stichproben gehen auf die selbe Grundgesamtheit zurück), indem getestet wird, ob die Frequenzen im einen Korpus normalerweise höher oder tiefer sind als im Vergleichskorpus. Die Summe der Ränge der kleineren Gruppe wird der Variable  $R_1$  zugewiesen, jene der größeren Gruppe  $R_2$ .<sup>10</sup> Die Berechnung von  $U$  erfolgt dann mit folgender Formel:

$$U_1 = N_1 N_2 + \frac{N_1(N_1 + 1)}{2} - R_1$$

$$U_2 = N_1 N_2 + \frac{N_2(N_2 + 1)}{2} - R_2$$
(7.10)

$N_1$  und  $N_2$  sind die Anzahl der Beobachtungen von Gruppe 1 und 2.  $U$  der kleinere Wert von  $U_1$  und  $U_2$ . Der kritische Wert für  $U$  kann

<sup>9</sup> Ich folge hier den Darstellungen in Kilgarriff (2001), Oakes (1998, 17) und Toutenburg (2000b, 174ff.).

<sup>10</sup> Bei gleich großen Gruppen, wie im dargestellten Fall, spielt die Verteilung auf die Variablen keine Rolle.

in einer entsprechenden Tabelle nachgeschlagen werden (z. B. Oakes 1998, 261). Für  $N_1 = 5$  und  $N_2 = 5$  muss  $U$  bei einem Signifikanzniveau von  $p \leq 0,05$  kleiner oder gleich 2 sein. Die in Tabelle 7.2 auf der vorherigen Seite wiedergegebene Verteilung wäre demnach nicht signifikant ( $U = 9$ ). Dies trotz des Ausschlags 88 in Korpus X, was genau dem Wunsch entspricht, dass Klumpen die Signifikanz nicht zu stark beeinflussen.

Treten bei der Rangliste der Frequenzen sog. Bindungen auf, tritt also eine Frequenz mehrfach auf, muss die  $U$ -Statistik korrigiert werden (vgl. Toutenburg 2000b, 176ff.). Es empfiehlt sich, hierzu ein Statistikprogramm zu benutzen.

Um im Vergleich von Korpora die jeweils typischen Wörter zu extrahieren, eignet sich nach Kilgarriff (2001, 127) der Mann-Whitney-Rank-Test besser als andere Testverfahren, wie z. B. der  $\chi^2$ -Test, der hochfrequente Wörter zu stark gewichtet. Allerdings ist die Berechnung im Vergleich zu anderen Tests, wie dem  $\chi^2$ -Test, relativ komplex.

### t-Test

Der t-Test prüft die Differenz zwischen beobachteten und erwarteten Werten auf Basis des arithmetischen Mittels und der Varianz zweier Stichproben (Manning/Schütze 2002, 163ff.). Um Kollokationen zu berechnen, wird das Korpus als Sequenz von  $N$  Bigrammen betrachtet (Manning/Schütze 2002, 164f.). Es werden die Wahrscheinlichkeiten berechnet, mit denen die Wörter  $w_1$  und  $w_2$  einzeln erscheinen und mit der Nullhypothese  $H_0$  die Unabhängigkeit der beiden Wörter behauptet.

Beim t-Test stellt sich ebenfalls das Problem der Verteilung sprachlicher Daten. Die Nullhypothese wird rasch verworfen, da der Test von einer Normalverteilung ausgeht. Dieser Test kann deshalb für die Berechnung von Kollokationen und auch für den Vergleich von Korpora nicht als Signifikanztest verwendet werden, sondern höchstens als Maß zur Erstellung einer Rangordnung von Korrelationen (Manning/Schütze 2002, 166, Kilgarriff 2001, 104).

### Exakter Test nach Fisher

Im Gegensatz zum t-Test und zum  $\chi^2$ -Test kann der Exakte Test nach Fisher auch für geringe Beobachtungsfrequenzen verwendet werden. Allerdings ist der Test rechenintensiv und seine Vorteile umstritten: Es ist unklar, ob sich die Resultate genügend stark von anderen Testverfahren, wie z. B. dem  $\chi^2$ -Test, unterscheiden (Manning/Schütze 2002, 189).

Pedersen (1996) schlägt diesen Test vor, um Bigramme zu berechnen. Damit kann das Problem von geringen und verzerrten Datenmengen verhindert werden. Der Test wird auch in Kilgarriff (2001, 105) erwähnt und von Stefanowitsch/Gries (2003, 218) zur Berechnung von Collocations angewandt.

### Pointwise Mutual Information

Dieses von der Informationstheorie motivierte Maß drückt den Informationsgehalt von  $x$  aus, wenn bekannt ist, dass auch  $y$  existiert (Manning/Schütze 2002, 178ff.). Für die Berechnung von Kollokationen eignet sich dieser Test vor allem bei geringen Fallzahlen jedoch kaum (Dunning 1993, 62; Manning/Schütze 2002, 179f.). Kilgarriff (2005, 266ff.) weist den Test aber wegen der fehlenden Normalverteilung von Sprache gerade auch für hohe Fallzahlen zurück.

## 7.3 Weitere Hilfsmittel

### 7.3.1 Dispersion: Juillands D

In Kapitel 7.2.1 wurde bereits angesprochen, dass Wörter in Klumpen verteilt auftreten können. Mit dem Mann-Whitney-Rank-Test steht ein Test zur Verfügung, der diese Klumpenhaftigkeit relativiert, indem das Korpus in mehrere Teile gesplittet wird (vgl. Kapitel 7.2.5). Um jedoch direkt in einem Maß auszudrücken, ob ein Wort oder eine Wortgruppe im Korpus eher gleichmäßig oder ungleichmäßig verteilt ist, können sog. Dispersionsmaße verwendet werden (Oakes 1998, 189ff.).

Drei verbreitete Maße sind Juillands  $D$ , Carrolls  $D_2$  und Rosengrens  $S$  (Oakes 1998, 189ff.), wobei ich hier nur auf ersteres eingehen möchte.

Teil	NZZ-Korpus Januar 1995				
	<i>Terror</i> (Ausland)	<i>Neujahr</i> (Ausland)	<i>Neujahr</i> (Inland)	<i>in der Nacht</i> (Vermisch- tes, Zürich)	<i>ums Leben</i> (Vermisch- tes, Zürich)
1	12	6	3	6	8
2	5	2	0	4	4
3	5	3	4	7	0
4	11	3	0	2	1
5	11	6	0	7	4
6	4	2	0	2	3
7	5	1	0	3	5
8	20	0	0	3	5
9	11	2	0	7	2
10	21	3	0	5	3
$\bar{x}$	10,5	2,8	0,7	4,6	3,5
s	5,8	1,83	1,42	1,96	2,16
V	0,55	0,65	2,03	0,43	0,62
D	0,82	0,78	0,32	0,86	0,79

**Tabelle 7.3:** Berechnung des Dispersionsmaßes Juillands D für einige Einzelwörter und Kollokationen im NZZ-Korpus ( $\bar{x}$ : Mittelwert, s: Standardabweichung, V: Variationskoeffizient, D: Dispersion).

Juillands D wird auf Basis des Variationskoeffizienten V berechnet (vgl. Kapitel 7.1):

$$D = 1 - \frac{V}{\sqrt{n-1}} \quad (7.11)$$

Der Wert D liegt jeweils zwischen 0 und 1, wobei 1 eine absolut gleichmäßige, 0 eine absolut ungleichmäßige Dispersion ausdrückt.

### Diskussion

Oakes (1998, 191) zitiert Lyne (1985), der die drei oben erwähnten Maße zur Berechnung der Dispersion empirisch miteinander verglich. Dabei erweist sich D als zuverlässigstes Maß.

Tabelle 7.3 zeigt Beispielberechnungen von D an Frequenzen von Einzelwörtern und Kollokationen im NZZ-Korpus. Das Korpus wurde jeweils in zehn etwa gleich viele Wörter zählende Teilkorpora aufgeteilt. Das Maß gibt intuitiv nachvollziehbar an, ob eine gleichmäßige oder eher ungleichmäßige Verteilung vorliegt und kann, im

Gegensatz zum Variationskoeffizienten, nur Werte zwischen 0 und 1 einnehmen.

### 7.3.2 Differenzkoeffizient

Um die Differenz der Frequenzen von Ausdrücken in Teilkorpora untereinander auszudrücken, kann der Differenzkoeffizienten  $D$  nach Belica (1996, 68) berechnet werden:

$$D = \frac{A(C + D) - C(A + B)}{A(C + D) + C(A + B)} \quad (7.12)$$

Die Werte von  $D$  liegen immer zwischen 1 und  $-1$  und lassen deshalb Vergleiche mit den Differenzkoeffizienten jeweils anderer Ausdrücke zu.

## 7.4 Fazit

Tabelle 7.4 auf der nächsten Seite gibt einen Überblick über die besprochenen Hypothesentests. Außer dem Mann-Whitney-Rank-Test sind alle Verfahren wegen der fehlenden Normalverteilung in sprachlichen Daten für Signifikanztests, in denen Wörter und Wortkombinationen gezählt werden, nur bedingt verwendbar. Um Ranglisten unterschiedlich starker Signifikanzen zu erstellen, eignen sich die Tests jedoch fast alle mehr oder weniger gut; Unterschiede gibt es in der Gewichtung von geringen bzw. hohen Frequenzen.<sup>11</sup> Werden die Signifikanztests jedoch für die Verteilung von Artikeln (in denen ein oder mehrere bestimmte Phänomene auf Wortebene existieren) verwendet, entschärft sich das Problem der Klumpenhaftigkeit sprachlicher Daten.

<sup>11</sup> Kilgarriff (2001, 113) hält es für unwahrscheinlich, dass es eine Statistik geben kann, die sich gleichzeitig für hohe und niedrige Frequenzen eignet: „Linguists have long made a distinction approximating to the high/low frequency contrast: form words (or ‘grammar words’ or ‘closed class words’) vs. content words (or ‘lexical words’ or ‘open class words’). The relation between the distinct linguistic behaviour, and the distinct statistical behaviour of high-frequency words is obvious yet intriguing. It would not be surprising if we cannot find a statistic which works well for both high and medium-to-low frequency words. It is far from clear what a comparison of the distinctiveness of a very common word and a rare word would mean.“

Statistik		Zweck		Kommentar
	Kollokationen berechnen	Korpusvergleich		
		Basis Wörter	Basis Artikel	
$\chi^2$	Ranglisten: ja; Signifikanztest: nein	Ranglisten: ja; Signifikanztest: nein	ja	Problem Normalverteilung, präzise für Korpusvergleiche
$G^2$	Ranglisten: ja; Signifikanz: nein	Ranglisten: ja; Signifikanztest: nein	ja	Problem Normalverteilung, robust bei geringen und hohen Frequenzen
U	ja	ja	ja	Lindert das Problem der fehlenden Normalverteilung; hohe Frequenzen werden weniger stark gewichtet; komplexes Verfahren
t-Test	Ranglisten: bedingt; Signifikanz: nein	Ranglisten: bedingt; Signifikanz: nein	ja	Problem Normalverteilung
Fisher	ja	ja	ja	Für geringe Frequenzen; Vorteile unklar; rechenintensiv
MI	nein	nein	nein	Problem Normalverteilung; überschätzt geringe Frequenzen

**Tabelle 7.4:** Statistische Maße und ihre Stärken und Schwächen hinsichtlich korpuslinguistischer Funktionen.



Kilgarriff (2005, 273) plädiert für den Verzicht von Signifikanztests und stattdessen dafür, die Datenmenge zu erhöhen; etwas, was zumindest technisch immer weniger ein Problem ist:

Hypothesis testing has been used to reach conclusions, where the difficulty in reaching a conclusion is caused by sparsity of data. But language data, in this age of information glut, is available in vast quantities. A better strategy will generally be to use more data[.] Then the difference between the motivated and the arbitrary will be evident without the use of compromised hypothesis testing. As Lord Rutherford put it: „If your experiment needs statistics, you ought to have done a better experiment.“ (Kilgarriff 2005, 273)

Möchte man trotzdem mit statistischen Tests arbeiten, ist es entscheidend, möglichst exakt zu definieren, was diese leisten sollen. Um Kollokationen zu berechnen, ist eine genaue Definition von ‚Kollokation‘ unumgänglich: In Kapitel 6 wurde auf die verschiedenen Definitionen verwiesen. Für die vorliegende Arbeit muss nicht der strenge phrasologische Maßstab an potenziellen Kollokationen angesetzt werden, indem Idiomatizität verlangt wird. Daher können die Bewertungen der Tests hinsichtlich der Stärke, Kollokationen im Korpus zu finden, etwas relativiert werden: Es ist gerade aus corpus-driven-Sicht interessant, nicht nur Kollokationen zu finden, bei denen aus lexikalischer Sicht ohnehin schon von Kollokationen gesprochen würde, sondern eben auch Kollokationen zu erkennen, die nur aufgrund ihrer überzufälligen Verteilung statistisch auffallen.

Für die Wahl des Tests ebenfalls relevant ist die Wahl des Frequenzausschnitts. Wenn das Interesse hauptsächlich bei Kollokationen mit geringen Fallzahlen liegt, muss der Test auch mit geringen Fallzahlen umgehen können; umgekehrt, wenn hochfrequente Phänomene untersucht werden sollen. Letztlich hängt diese Frage davon ab, ob für das Untersuchungsinteresse eher strukturelle (und damit frequente) oder aber inhaltliche (und eher niedrigfrequente) Phänomene relevant sind (vgl. auch das Zitat in Fussnote 11 auf Seite 144).

In der vorliegenden Arbeit werde ich hauptsächlich mit folgenden Tests arbeiten: Für die Berechnung der signifikanten Kollokationen bzw. Mehrworteinheiten verwende ich den Log-Likelihood-Test, ebenso um Ranglisten zu erstellen und die Verteilung der Mehrworteinheiten in unterschiedlichen Korpora zu testen. Korpusvergleiche stelle ich größtenteils auf Basis der Artikelfrequenzen her und ver-

wende daher als statistisches Maß den  $\chi^2$ -Test, der für diese Zwecke geeignet ist und einfach berechnet werden kann.



## 8 Von corpus-driven zu corpus-based

Die Heuristik, wie sie in Kapitel 5.2 beschrieben wurde, gliedert sich in corpus-driven und corpus-based-Abschnitte. Während mit der Berechnung von statistisch signifikanten Kookkurrenzen versucht wird, musterhafte Strukturen im Text sichtbar zu machen (Kapitel 6), folgen nun die Schritte der Datenstrukturierung und Dateninterpretation, um zu einer Diskursbeschreibung zu gelangen. Damit ist das Scharnier zwischen corpus-driven und corpus-based erreicht. Besonders bei interpretativen Schritten ist es noch weniger möglich, theorieelos die Daten zu analysieren. Allerdings müssen auch diese Schritte stark korpusgeleitet vollzogen werden, aber eben eher als corpus-based statt als corpus-driven-Analyse.

Die Interpretation gliedert sich in drei Schritte:

1. Durch Sortierung, Filterung und Kontrastierung der Mehrworteinheiten wird die Datenmenge reduziert (Kapitel 8.1).
2. Die eigentliche Interpretation der Mehrworteinheiten geschieht unter bestimmten Aspekten, die nach der Typik der Muster fragen (Kapitel 8.2).
3. Die Heuristik ist zirkulär angelegt. Deshalb erfolgt daraufhin ein Schritt zurück in die Daten, der die gewonnenen Erkenntnisse überprüft und den Prozess neu starten kann (Kapitel 8.3).

### 8.1 Typikprofile erstellen

Die besondere Herausforderung bei einer corpus-driven-Analyse liegt darin, potenziell interessante von restlichen Daten zu scheiden. Denn mit den Berechnungen entstehen umfangreiche Daten. Der Datenumfang hängt von den Parametern ab, die für die Berechnung der Mehrworteinheiten gewählt wurden. Es würde der Philosophie des corpus-driven-Paradigmas jedoch widersprechen, wenn bereits dort

besonders rigide Parameter gewählt würden. Deshalb muss der Schritt des Filterns der Daten möglichst spät erfolgen.

Es gibt eine Reihe von Möglichkeiten, die Datenmenge zu verkleinern. Im Grundsatz erfolgt dies durch folgende Methoden:

1. Sortieren: Die Listen der berechneten Mehrworteinheiten können, abhängig von der Berechnung, nach unterschiedlichen Kriterien sortiert werden. Immer verfügbar ist das Kriterium der Frequenz. Wenn Mehrworteinheiten für unterschiedliche Teilkorpora berechnet worden sind, ist es sinnvoll, neben der absoluten Frequenz auch eine relative Frequenz zu berechnen, die abhängig ist von einer für alle Teilkorpora gleichermaßen gültigen Referenzgröße („Normalisierung“). Die einfachste Größe ist die Anzahl der Wörter im gesamten Korpus oder im Teilkorpus. Sind bei der Berechnung der Mehrworteinheiten statistische Maße angewandt worden, können die Listen auch danach sortiert werden. Weitere sinnvolle Kriterien sind die Länge der Mehrworteinheiten oder bei diskontinuierlichen Mehrworteinheiten die Varianz der Distanz zwischen dem ersten und letzten Token des Ausdrucks (vgl. Formel 7.2 auf Seite 133).  
Die Sortierung alleine verkleinert die Datenmenge noch nicht; erst kombiniert mit der Filterung wird dies möglich.
2. Filtern: Für die Filterung der Daten stehen wieder eine Reihe von Kriterien zur Verfügung, deren Wahl stark vom Untersuchungsinteresse abhängt. Denkbar ist die Konzentration auf Mehrworteinheiten, die
  - bestimmte Wörter oder Lemmata enthalten,
  - bestimmte morphologische oder syntaktische Eigenschaften aufweisen,
  - eine bestimmte Mindestlänge,
  - eine Mindestfrequenz oder
  - eine Mindestsignifikanz aufweisen.
3. Kontrastieren: Wenn das Untersuchungskorpus entlang diachroner oder synchroner Achsen in mehrere Teilkorpora unterteilt

ist, können die Mehrworteinheiten separat für jedes Teilkorpus berechnet werden. Anschließend ist es sinnvoll, die Listen der Teilkorpora gegeneinander zu kontrastieren, wobei folgende Teilmengen der Mehrworteinheiten entstehen:<sup>1</sup>

- Vereinigungsmenge: Die Menge alle Mehrworteinheiten aller Teilkorpora; sie ist interessant, um Aussagen unabhängig von der Einteilung in Teilkorpora zu machen.
- Schnittmenge: Die Menge aller Mehrworteinheiten, die in allen Teilkorpora vorkommen. Hierbei handelt es sich hauptsächlich um Mehrworteinheiten, die typisch für das ganze Korpus sind.
- Differenzmengen: Alle Mehrworteinheiten, die jeweils nur in einem der Teilkorpora vorkommen. Hierbei handelt es sich um Mehrworteinheiten, die typisch für das Teilkorpus sind und dieses von den anderen Teilkorpora unterscheiden.

Diese Mengen gehen von simplen Ja/Nein-Entscheidungen aus: Eine Kollokation kommt entweder nur im einen, nur im anderen oder in beiden Korpora vor. Statistische Signifikanztests erlauben aber eine differenziertere Sichtweise: Sie setzen die Frequenzen der Mehrworteinheiten in Relation zu den Korpusgrößen.

Die Verfahren des Sortierens und Filterns bedürfen keiner weiteren methodischen Erklärungen.<sup>2</sup> Für die Kontrastierung und Bildung von Vereinigungs-, Schnitt- und Differenzmengen sind ebenfalls keine komplexeren Verfahren nötig. Anders sieht es für statistische Kontrastierungen aus. In Kapitel 7 wurden einige Möglichkeiten des statistischen Korpusvergleichs aufgrund der Distribution von Wörtern oder Kollokationen aufgezeigt. Grundsätzlich werden dabei für alle oder

---

<sup>1</sup> In Bezug auf Kollokationen spricht bereits Firth (1957, 196) von zwei für sprachstilistische Untersuchungen interessanten Typen: „[...] a large number of collocations which have been common property for long periods and are still current even in everyday colloquial. This method of approach makes two branches of stylistics stand out more clearly: (a) the stylistics of what persists in and through change, and (b) the stylistics of personal idiosyncrasies.“

<sup>2</sup> Ein Tabellenkalkulationsprogramm, eine Datenbank oder die Standard-Unix-Befehle ermöglichen diese Operationen.

einen Teil der Mehrworteinheiten Kontingenztabellen erstellt, wie in Kapitel 7.2.2 dargestellt.

Danach kann mit einem Signifikanztest geprüft werden, ob diese Verteilung überzufällig ist (die Nullhypothese verworfen werden kann). Jedoch ist deutlich geworden, dass die meisten Statistiken, die von einer Normalverteilung der Daten ausgehen, zu rasch die Nullhypothese verwerfen und eine signifikante Verteilung sehen (vgl. Kapitel 7.2.1). Daher muss entweder ein anderes Testverfahren verwendet werden, oder aber der statistische Test soll nicht über die Signifikanz entscheiden, sondern nur über die Rangordnung bezüglich der Stärke der Korrelation zwischen Kollokation und Teilkorpus.

Als bestes statistisches Verfahren, um die ungleichmäßige Verteilung im Teilkorpus auszugleichen, hat sich der Mann-Whitney-Rank-Test bewährt (vgl. Kapitel 7.2.5). Der Log-Likelihood  $G^2$ -Test geht zwar von einer Normalverteilung aus, ist jedoch vor allem im Vergleich zum  $\chi^2$ -Test robuster bei geringen Frequenzen und wird ebenfalls in einer Reihe von Arbeiten für Korpusvergleiche verwendet (vgl. Kapitel 7.2.4).

Durch die Kontrastierung der Listen von Mehrworteinheiten für bestimmte Teilkorpora werden aus diesen Listen die für die jeweiligen Korpora typischen Mehrworteinheiten extrahiert. Diese Zusammenstellungen von für ein Teilkorpus typischen Mehrworteinheiten (im Kontrast zu einem oder mehreren anderen Teilkorpora) nenne ich ‚Typikprofil‘. Natürlich können weitere Kontrastierungen mit Typikprofilen anderer Korpora folgen. So kann es interessant sein, das Typikprofil von Artikeln des Ressorts ‚Ausland‘ der NZZ mit einem Typikprofil von ähnlichen Artikeln in einer anderen Zeitung zu vergleichen. Oder es kann für einen beliebigen Text berechnet werden, wie stark er mit einem bestimmten Typikprofil übereinstimmt: Je mehr Mehrworteinheiten eines Typikprofils in einem Text vorkommen, desto eher entspricht der Text dem Typikprofil.

## 8.2 Interpretieren: Diskurse modellieren

Der Schritt der Interpretation hat zum Ziel, die gewonnenen Daten musterhafter Strukturen in den Fokus des Untersuchungsinteresses

zu stellen. Im Hinblick auf eine Modellierung sind grundsätzlich zwei Vorgänge denkbar:

- Abstrahierung: Die Beobachtung von musterhaften Strukturen macht es möglich, diese zu einem allgemeineren Phänomen zu abstrahieren.
- Konkretisierung: Das Entdecken von musterhaften Strukturen kann der Ausgangspunkt sein, um die Verwendung und die Gestalt des Musters konkreter zu beschreiben.

Unter diesen beiden Vorgängen gibt es nun eine Vielzahl von Aspekten, denen nachgegangen werden kann. Ich versuche dieses weite Feld der Interpretationsmöglichkeiten etwas zu systematisieren (vgl. Abbildung 8.1 auf der nächsten Seite): Die Interpretation beginnt bereits durch die Einnahme bestimmter Perspektiven auf die Daten. Ich plädierte für eine Perspektive, die die Sprechweisen im Korpus ins Blickfeld nimmt. Daneben soll die Frage nach den Inhalten jedoch nicht völlig vernachlässigt werden. Doch entscheidend ist, dass nach der Typik im Korpus gefragt wird. Die Fragen lauten also: Welche typischen Sprechweisen und welche typischen Sprechinhalte sind im Korpus zu finden?

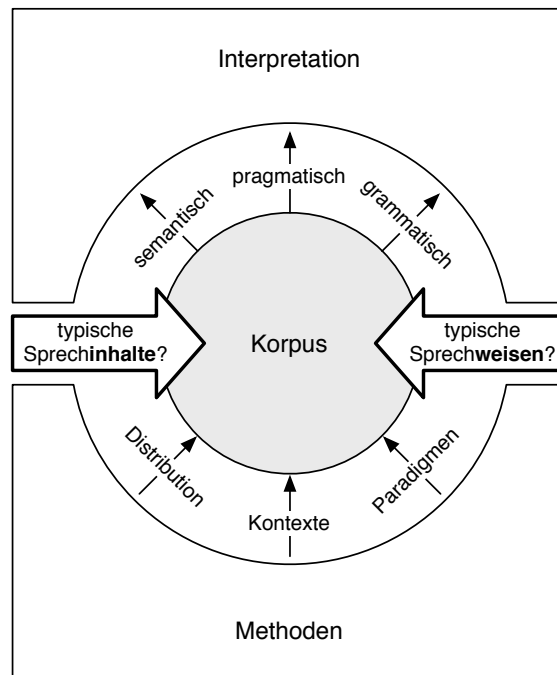
Weiter werden für die interpretative Beantwortung dieser Fragen Methoden verwendet, die ich in drei Bereichen angesiedelt sehe: Distribution von Elementen, Kontexte der Elemente und semantische Paradigmen aus Elementen. Diese Methoden werden in den folgenden Unterkapiteln ausgeführt.

Die Interpretation vollzieht sich letztlich ebenfalls vor unterschiedlichen Hintergründen. Die Beobachtungen können semantisch, pragmatisch oder eher grammatisch interpretiert werden. Gängige Aspekte diskursanalytischer Arbeiten wären z. B. Fragen nach der Verwendung bestimmter Begriffe, Metaphern oder Argumente, die im semantischen, pragmatischen oder grammatischen Licht betrachtet werden.<sup>3</sup> Es versteht sich von selbst, dass in der konkreten Analyse diese Abgrenzungen kaum trennscharf gehalten werden können; sie dienen bloß als Richtschnur, an denen die interpretativen Schritte ausgerichtet werden können.

---

<sup>3</sup> Vgl. zu diesen gängigen Aspekten Jung (1996, 464f.).





**Abbildung 8.1:** Perspektiven, Methoden und Hintergründe der Interpretation.

Das Ziel des interpretativen Schrittes ist eine Diskursbeschreibung. Die Art der Beschreibung hängt vom Untersuchungsinteresse ab. Im vorliegenden Fall handelt es sich um eine Beschreibung, die die Sprechweisen, d. h. den Sprachgebrauch in der NZZ zwischen 1995 und 2005 umfasst. Aus der Analyse dieser Sprechweisen können Hinweise auf Diskurse gefunden werden. Zudem werden diese Diskurse über ihre für sie je typischen Sprechweisen beschrieben. Die Beschreibungen sollen also eine Menge von Beobachtungen und Thesen über die Art dieser Diskurse enthalten, wobei die Balance zwischen konkreter Beschreibung und Abstrahierung gefunden werden muss. Denn die Beschreibungen sollen gleichzeitig zusammenfassen und verallgemeinern und trotzdem dem Untersuchungsgegenstand beschreibungsadäquat sein. In der vorliegenden Untersuchung werde ich nur für einen Bruch-

Distribution	Kontexte	Paradigmen
Grad der Typik; Typikprofile und deren Vergleiche	Kontextualisierungsprofi- le; pragmatische Funktionen	paradigmatische Konstel- lationen

**Tabelle 8.1:** Methoden der Interpretation

teil der gefundenen Muster zu Sprechweisen abstrahieren und dadurch bloß erste Hypothesen für die Art der gefundenen Diskurse bilden können.

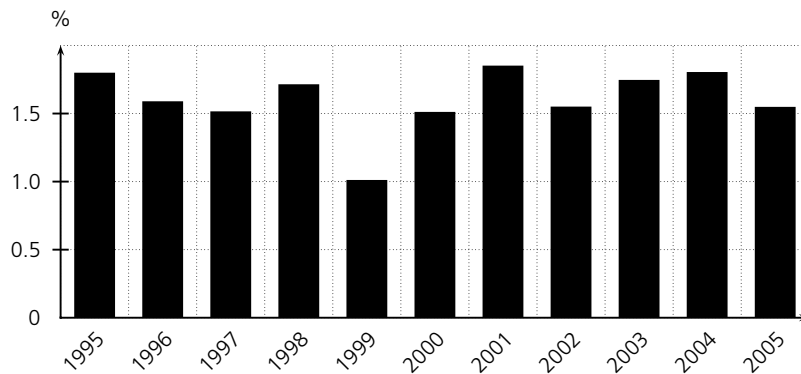
Wie oben angekündigt entwickle ich im Folgenden vier Methoden der Interpretation. Zusammen mit der bereits oben in Kapitel 8.1 dargestellten Erarbeitung von Typikprofilen ergeben sich fünf Methoden, die in Tabelle 8.1 den drei Bereichen Distribution, Kontexte und Paradigmen zugeordnet sind.

### 8.2.1 Grad der Typik

Der Grad der Typik erfasst, wie typisch ein Muster für einen beliebigen Korpusausschnitt ist. Das einfachste Maß dafür ist die Frequenz des Musters: Wie oft tritt es in welchen Korpusabschnitten auf?

Der Ausdruck *Kampf gegen Terrorismus* ist eine für die Periode von 2003–2005 im NZZ-Korpus typische Mehrworteinheit (vgl. Kapitel 13.1.1). An diesem Beispiel möchte ich aufzeigen, was mit ‚Grad der Typik‘ gemeint ist. Abbildung 8.2 auf der nächsten Seite zeigt die zur jeweiligen Korpusgröße relativen Frequenzen von Artikeln im NZZ-Korpus, die den verkürzten Ausdruck *Kampf gegen* enthalten. In den Daten sind Schwankungen ersichtlich. Es stellt sich jedoch die Frage, ob die beobachteten Unterschiede zwischen den Teilkorpora statistisch signifikant sind. Es bieten sich eine Reihe von statistischen Tests an, um hier Sicherheit zu gewinnen (vgl. dazu Kapitel 7). Da sich der  $\chi^2$ -Test sehr einfach berechnen lässt und für Korpusvergleiche auf Basis von Artikeln zuverlässig ist, werde ich primär mit diesem Test arbeiten.<sup>4</sup> Das Verfahren gestaltet sich also wie folgt (vgl. für Details Kapitel 7.2.3):

<sup>4</sup> Belica (1996) z. B. verwendet für einen ähnlichen Zweck ebenfalls diesen Test.



**Abbildung 8.2:** Die relativen Frequenzen in Prozent von Artikeln im NZZ-Korpus, die den Ausdruck *Kampf gegen* enthalten. Grundgesamtheit: 44 843 Artikel; Details vgl. Kapitel 10.

Zunächst müssen die absoluten Frequenzen (O) in eine Kontingenztafel eingetragen und die zu erwartenden Frequenzen (E) berechnet werden. Die Kontingenztafel für das Beispiel *Kampf gegen* im NZZ-Korpus ist als Tabelle 8.2 auf der nächsten Seite dargestellt.<sup>5</sup>

Nun kann für die Kontingenztafel  $\chi^2$  berechnet werden. Für die Frequenzen von *Kampf gegen* beträgt dieser Wert 15,635. Um die Signifikanz von  $\chi^2$  zu bestimmen, muss geprüft werden, mit welcher Wahrscheinlichkeit p die Nullhypothese  $H_0$  verworfen werden kann. Die Nullhypothese lautet: Die beobachteten Frequenzen sind nicht genügend unterschiedlich, um den Unterschied nicht durch Zufall erklären zu können.

Wenn  $H_0$  mit 95 %iger Sicherheit ausgeschlossen werden soll (Signifikanzniveau von 0,05), muss bei einem Freiheitsgrad<sup>6</sup> von 10  $\chi^2 \geq 18,31$  sein, bei Signifikanzniveaus von 0,01 und 0,001 muss

<sup>5</sup> In den kommenden Kontingenztafeln werde ich der Übersichtlichkeit wegen auf die Darstellung der erwarteten Werte verzichten. Weitere Informationen zur Berechnung finden sich in Kapitel 7.2.3.

<sup>6</sup> Der Freiheitsgrad df einer Kontingenztafel berechnet sich wie folgt: (Reihenzahl – 1) × (Spaltenzahl – 1). Bei einer 2 × 11-Tabelle beträgt er also 10.

Jahr	<i>Kampf gegen</i>		<i>¬Kampf gegen</i>		Total
	O	E	O	E	
1995	72	61,48	3594	3604,52	3666
1996	61	62,23	3650	3648,77	3711
1997	59	64,18	3768	3762,82	3827
1998	63	60,64	3553	3555,36	3616
1999	46	71,24	4202	4176,76	4248
2000	71	75,40	4425	4420,6	4496
2001	85	73,33	4288	4299,67	4373
2002	72	73,53	4313	4311,47	4385
2003	76	71,02	4159	4163,98	4235
2004	80	69,68	4075	4085,32	4155
2005	67	69,28	4064	4061,72	4131
Total	718		44125		44843
$\chi^2 = 15,635$ , $df = 10$ , $p > 0,10$ (nicht signifikant)					
Cramér's $V = 0,019$ , $\phi = 0,019$					

**Tabelle 8.2:** Kontingenztafeln für *Kampf gegen* (alle Flexionsformen) im NZZ-Korpus (Basis: Artikel).

$\chi^2 \geq 23,21$  bzw. 29,59 sein.<sup>7</sup> Der oben berechnete Wert von  $\chi^2 = 15,635$  ist somit nicht signifikant; es gibt eine zu schwache Korrelation zwischen Jahr und Frequenz. Die Veränderungen der Frequenzen sind höchstwahrscheinlich zufällig entstanden.

In Tabelle 8.2 sind zusätzlich die Werte für Cramér's  $V$  und  $\phi$  angegeben, die die Stärke der Beziehung ausdrücken (Details dazu finden sich in Kapitel 7.2.3). Diese Werte sind sehr niedrig; es besteht also eine sehr schwache Korrelation zwischen Jahr und Frequenzen, was nicht überraschend ist, da die Korrelation nicht signifikant ist.<sup>8</sup>

Wird der Ausdruck *Kampf gegen* erweitert auf *Kampf gegen den Terror*, ändert sich das Bild, wie die Abbildung 8.3 auf der nächsten Seite und die Werte in Tabelle 8.3 auf der nächsten Seite zeigen.

Die Statistik zeigt, dass zwar eine hochsignifikante Korrelation zwischen Jahr und Frequenz besteht (man kann mit mindestens 99,9 % davon ausgehen, dass eine Abhängigkeit zwischen Jahr und Frequenz

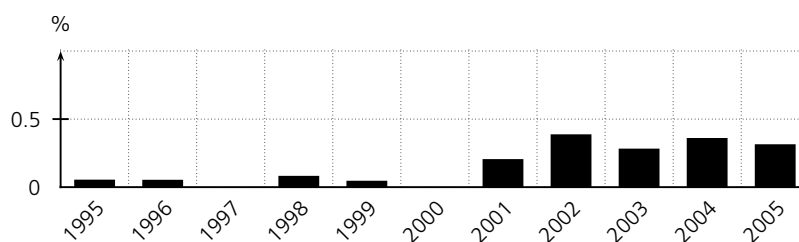
<sup>7</sup> Diese Werte sind einer entsprechenden Tabelle der Signifikanzniveaus von  $\chi^2$  zu entnehmen, wie sie in den meisten Statistikbüchern abgedruckt sind, so z. B. in Albert/Koster (2002, 179).

<sup>8</sup> In meinen Berechnungen unterscheiden sich Cramér's  $V$  und  $\phi$  kaum. Zudem sind die Werte meist sehr klein. Ich verzichte deshalb im Folgenden auf die Angabe von  $\phi$  und erwähne Cramér's  $V$  nur im Ausnahmefall.

Jahr	<i>Kampf gegen (den) Terror</i>	$\neg$ <i>Kampf gegen (den) Terror</i>	Total
1995	2 (0,05 %)	3664	3666
1996	2 (0,05 %)	3709	3711
1997	0 (0 %)	3827	3827
1998	3 (0,08 %)	3613	3616
1999	2 (0,05 %)	4246	4248
2000	0 (0 %)	4496	4496
2001	9 (0,21 %)	4364	4373
2002	17 (0,39 %)	4368	4385
2003	12 (0,28 %)	4223	4235
2004	15 (0,36 %)	4140	4155
2005	13 (0,31 %)	4118	4131
Total	75 (0,17 %)	44768	44843

$\chi^2 = 56,095$ ,  $df = 10$ ,  $p < 0,001$  (signifikant)  
 Cramér's  $V = 0,035$ ,  $\phi = 0,035$

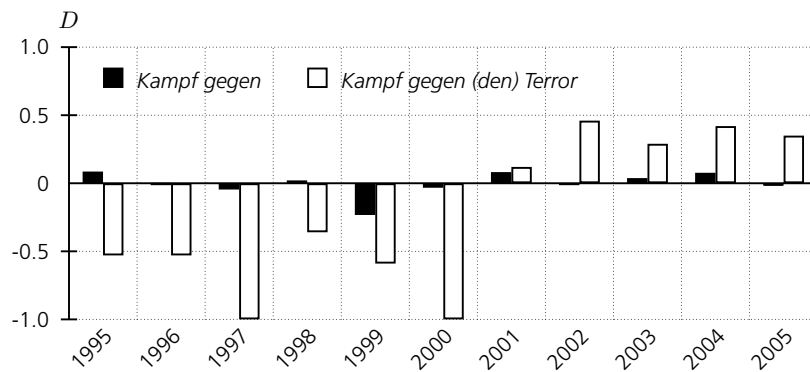
**Tabelle 8.3:** Statistik von *Kampf gegen (den) Terror* (inkl. Flexionsformen) im NZZ-Korpus (Basis: Artikel). In Klammern: Frequenzen in Prozent zum Total.



**Abbildung 8.3:** Die relativen Frequenzen in Prozent von Artikeln im NZZ-Korpus, die den Ausdruck *Kampf gegen (den) Terror* (inkl. Flexionsformen) enthalten.

existiert), allerdings ist die Beziehung schwach. D. h., die Unterschiede der Verteilung über die Jahre sind signifikant, weil die Stichprobe sehr groß ist, die Beziehung ist aber schwach, da es insgesamt sehr wenige Fälle gibt, in denen der Ausdruck *Kampf gegen (den) Terror* auftritt.

Die Unterschiede zwischen den beiden Ausdrücken *Kampf gegen* und *Kampf gegen (den) Terror* treten noch deutlicher hervor, wenn die Verläufe mit einem Differenzkoeffizienten dargestellt werden (vgl. zu den mathematischen Details Kapitel 7.3.2). Tabelle 8.4 zeigt für die



**Abbildung 8.4:** Die Differenzkoeffizienten D für die Ausdrücke *Kampf gegen* und *Kampf gegen (den) Terror* (inkl. Flexionsformen) im NZZ-Korpus (Basis Artikel).

zwei Ausdrücke den Koeffizienten D, der die jeweilige Abweichung zum Mittel ausdrückt.

Alternativ können diese Berechnungen auch auf Basis der Wörter statt der Artikel vorgenommen werden. In Kapitel 7.2.1 habe ich das Problem der Klumpen im Sprachgebrauch aufgezeigt und deshalb vorgeschlagen, Frequenzberechnungen primär auf Basis von Artikeln vorzunehmen, jedoch Vergleiche mit Berechnungen auf Basis der Wörter vorzunehmen.

Im Fall der Berechnungen oben zeigt sich das Bild wie folgt: *Kampf gegen* zeigt zwar eine geringe Signifikanz ( $\chi^2 = 22,783$ ,  $df = 10$ ,  $p < 0,025$ ), allerdings ist die Korrelation sehr schwach (Cramér's  $V = 0,001$ ). Das bedeutet, dass in einzelnen Jahrgängen zwar mehr Artikel vorkommen, die den Ausdruck enthalten, in diesen jedoch auch eine Häufung des Ausdrucks vorkommt, so dass die Korrelation auf Basis der Wörter leicht signifikant ist. Allerdings ist dies aufgrund der Klumpenbildung (vgl. Kapitel 7.2.1) keine Überraschung.

Im Fall von *Kampf gegen den Terror* zeigt sich ebenfalls eine signifikante Korrelation ( $\chi^2 = 59,418$ ,  $df = 10$ ,  $p < 0,001$ ; Cramér's  $V = 0,001$ ), die etwas schwächer ist, als bei der Berechnung auf Basis von Artikeln.

Trotz dieser leichten Differenzen ist also offensichtlich: Die Typik der beiden Ausdrücke *Kampf gegen* und *Kampf gegen den Terror* charakterisiert sich unterschiedlich:

- *Kampf gegen* ist ein Ausdruck, der für das ganze Korpus typisch ist und der durchschnittlich in 1,68 % der Artikel zu finden ist, was vergleichsweise oft ist. Ich bezeichne diese Typik eher *hohen Grades* als *stabil*.
- *Kampf gegen (den) Terror* ist dagegen ein Ausdruck, der nur für einen Teil des Korpus typisch ist. Zwar liegt die höchste Jahresfrequenz bloß bei 0,39 % der Artikel im Jahr 2002; charakteristisch an diesem Ausdruck ist jedoch der hohe  $\chi^2$ -Wert. Ich bezeichne diese Typik *eher niedrigen Grades* als *variabel*.

Damit steht das Werkzeug zur Verfügung, um die Typik beliebiger Ausdrücke in beliebigen Teilen des Korpus zu beschreiben.

### 8.2.2 Kontextualisierungsprofile

Die Beispiele der Ausdrücke *Kampf gegen* und *Kampf gegen den Terror* im vorherigen Kapitel 8.2.1 haben gezeigt, dass es interessant sein kann, die typischen Kontexte von Ausdrücken genauer zu untersuchen. Wenn die Beobachtung gemacht werden kann, dass *Kampf gegen* ein stabiler Ausdruck hoher Typik ist, die erweiterte Form *Kampf gegen den Terror* aber variabel, muss der Fokus auf andere mögliche Kontexte von *Kampf gegen* gerichtet werden. Ich möchte dies als Suche nach ‚Kontextualisierungsprofilen‘ bezeichnen.

Für das obige Beispiel ergibt eine Analyse des Korpus die in Abbildung 8.5 auf Seite 162 dargestellten Kontextualisierungsprofile. Die Grafik zeigt die häufigsten im Syntagma rechts des Ausdrucks *Kampf gegen* stehenden Kollokatoren (in Relation zu allen Kollokationen), separiert durch den Zeitpunkt des 11. September 2001. Um diese Grafik zu erstellen, wurden alle Belege (842) von *Kampf gegen* im Korpus untersucht und automatisch die jeweils rechten Kollokatoren des Suchausdrucks extrahiert. Dabei gilt als Kollokator alles Wortmaterial bis und mit dem ersten groß geschriebenen Wort.<sup>9</sup> Damit wird sowohl

<sup>9</sup> Es wurde mittels regulärer Ausdrücke gesucht. Das entsprechende Suchmuster lautete: `(.+? [[:upper:]] .+?\s)`.

(*Kampf gegen das*) *organisierte Verbrechen* als auch (*Kampf gegen den*) *Terror* gefunden. Es wurden nur die Kollokatoren berücksichtigt, die in der Periode vor oder nach dem 11. September 2001 mindestens dreimal erschienen. Teilweise wurden Kollokatoren in unterschiedlichen Schreibvarianten, Flexionsformen oder mit Ergänzungen zu Gruppen zusammengefasst. Solche Zusammenfassungen sind durch die Verwendung des Sterns (\*) ersichtlich, der für beliebig viele Zeichen steht.

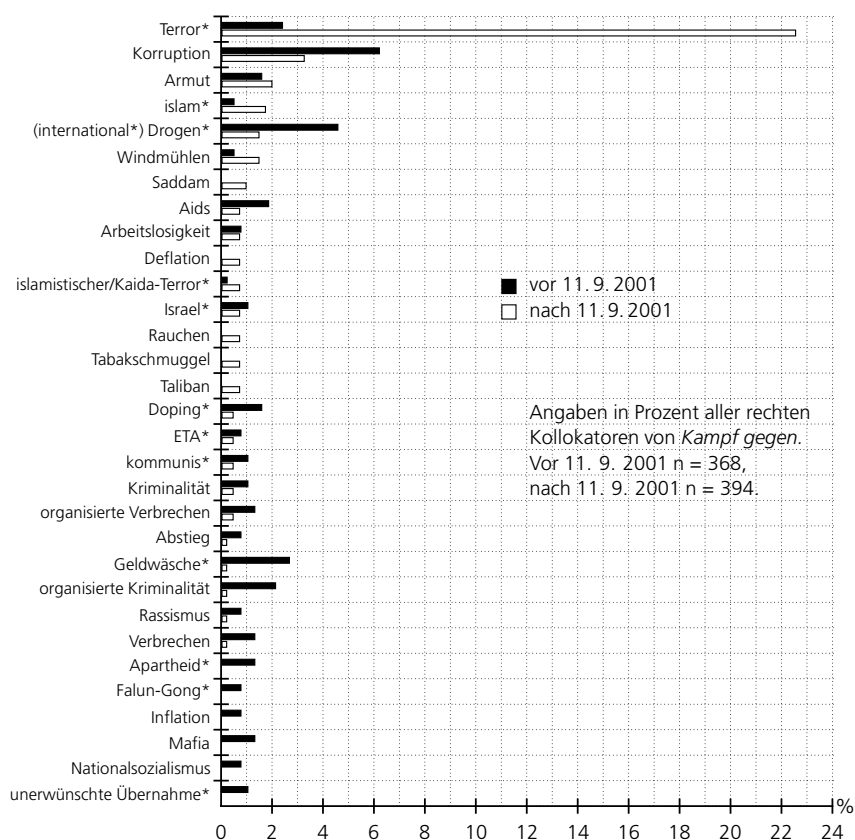
Die Übersicht der rechten Kollokatoren zeigt die inhaltlichen Gewichtungen, die für die beiden Zeitperioden typisch sind. Während in der ersten Zeitperiode der *Kampf* der *Korruption*, den *Drogen*, der *Geldwäsche* und der *Kriminalität* galt, wird das Feld in der zweiten Periode von *Terror* dominiert, gefolgt von *Korruption*, *Armut* und Wörter mit dem Morphem *islam*-.

Je nach Untersuchungsinteresse kann der Fokus auf bestimmten Kollokationen liegen oder aber möglichst breit sein. Es kann sinnvoll sein, Kollokationen zu Klassen, die z. B. semantisch definiert werden, zusammenzufassen, beispielsweise zu einer semantischen Klasse KRIEGERISCH.

Es gibt mehrere Möglichkeiten ein Kontextualisierungsprofil zu erstellen. Neben manuell oder halbautomatisch erstellten Profilen, bei denen die Kontexte eines bestimmten Ausdrucks ausgezählt, interpretiert und klassifiziert werden, können auch hier statistische Verfahren weiter helfen. Nicht nur dem Namen nach, auch inhaltlich gibt es Berührungspunkte zwischen dem Konzept der Kontextualisierungsprofile und den Kookkurrenzprofilen, wie sie Belica/Steyer (2006, 10) definieren (vgl. Seite 119). Letztere sind streng statistisch definiert als Kookkurrenzcluster, können aber die Basis für die Erstellung eines Kontextualisierungsprofils bilden. Denn ein Kookkurrenzprofil stellt die statistisch signifikanten Kookkurrenzcluster zu einem bestimmten Ausdruck dar. Theoretisch erlaubt dieses Konzept auch die Erstellung von Kookkurrenzprofilen zu komplexen Ausdrücken, die aus mehreren Wörtern bestehen.<sup>10</sup> Die Berechnung solcher Profile nach

<sup>10</sup> Es ist noch nicht möglich, solche Kookkurrenzprofile auf Basis komplexer Ausdrücke im COSMAS II-System des IDS und der Kookkurrenzdatenbank (Belica 2001–2006) zu berechnen. In naher Zukunft wird die Berechnung solcher Profile jedoch implementiert (persönliche Mitteilung von Cyril Belica an den Autor, Januar 2006).





**Abbildung 8.5:** Die Füllungen von X im Muster KAMPF GEGEN [ARTIKEL] X mit Mindestfrequenz 3 für die Zeit vor und nach dem 11. September 2001. „\*“ steht als Platzhalter für weitere Zeichen. Details zur Berechnung können dem Text entnommen werden.

Vorbild des IDS ist aber komplex und für eigene Korpora eher selten praktikabel.

Doch das Grundprinzip, zu einem bestimmten Ausdruck weitere statistisch signifikante Kollokatoren zu berechnen, kann auch mit einfacheren Mitteln umgesetzt werden. Prinzipiell können die selben Verfahren angewendet werden, wie sie in Kapitel 6 bereits beschrieben sind.

### 8.2.3 Paradigmatische Konstellationen

Als ‚paradigmatische Konstellation‘ bezeichne ich ein Paradigma, das aus komplexen Ausdrücken besteht, die semantisch ähnlich sind. So kann z. B. ein Paradigma MITTEILEN definiert werden, das aus den Ausdrücken *wurde mitgeteilt*, *nach Mitteilung*, *nach Angaben der*, *Bekanntgabe* etc. besteht. Die Definition und Zusammensetzung des Paradigmas richtet sich allein nach dem Untersuchungsinteresse und kann je nach Zweck weiter oder enger gefasst werden.

Die paradigmatische Konstellation versucht den Stilbegriff zu operationalisieren (vgl. Kapitel 4.1). Denn sie gibt aus einem grundsätzlich möglichen Paradigma von Ausdrucksvarianten an, welche Variante in einem bestimmten Sprachausschnitt tatsächlich gewählt wird.

Ausgangspunkt ist z. B. ein Muster, das aufgrund seiner variablen Typik auffällt. Es kann interessant sein, nach den alternativen Ausdrücken zu fragen, die an Stelle des Musters potenziell auftreten könnten. Die paradigmatische Konstellation bietet dann eine Übersicht über die Gestalt des Paradigmas an Ausdrucksvarianten, also über die Distribution der Ausdrücke im Korpus.

Beispielsweise ist in der Auslandberichterstattung im NZZ-Korpus der Ausdruck *in den Tod* durch eine variable Typik charakterisiert, wie die Tabelle 8.4 auf der nächsten Seite zeigt. Eine Durchsicht der Belege zeigt, dass sich dahinter meistens das Muster *in den Tod gerissen* bzw. *riss ... in den Tod* verbirgt.

Es handelt sich bei diesem Muster um eine von vielen Möglichkeiten, den Vorgang des Sterbens zu beschreiben. Denkbar sind alternative Formulierungen wie *wurden getötet*, *starben*, *wurden Opfer von* etc. Wenn die Zunahme eines bestimmten Musters des Paradigmas beobachtet werden kann, gibt es mindestens zwei Faktoren, die das erklären können. Daraus ergeben sich folgende Thesen:

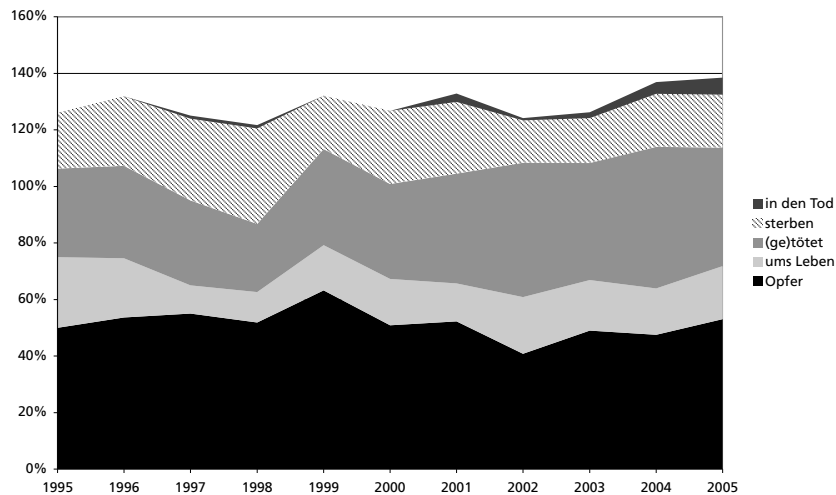
Jahr	<i>in den Tod</i>	<i>–in den Tod</i>	Total
1995	0 (0 %)	496	496
1996	0 (0 %)	493	493
1997	1 (0,2 %)	497	498
1998	1 (0,2 %)	503	504
1999	0 (0 %)	565	565
2000	0 (0 %)	592	592
2001	4 (0,7 %)	566	570
2002	1 (0,2 %)	508	509
2003	3 (0,53 %)	559	562
2004	5 (0,98 %)	503	508
2005	7 (1,44 %)	479	486
Total	22	5761	5783
$\chi^2 = 30,711$ , $df = 10$ , $p < 0,001$ (signifikant)			
Cramér's $V = 0,073$			

**Tabelle 8.4:**  $\chi^2$ -Statistik (Einheit: Artikel) für *in den Tod* in Artikeln des Ressorts Ausland im NZZ-Korpus. In Klammern: Frequenzen in Prozent zum Total.

1. Ereignisse, in denen potenziell der Ausdruck *in den Tod* verwendet werden könnte, haben zugenommen.
2. Der Ausdruck *in den Tod* wurde bezüglich oder auf Kosten anderer möglichen Ausdrücke des Paradigmas vermehrt verwendet.

Um zu überprüfen, ob eher These 1 oder These 2 gilt, müssen die paradigmatischen Konstellationen aller Fälle beschrieben werden, in denen eine oder mehrere Ausdrucksvarianten des Paradigmas überhaupt möglich wären. Es müssen also alle Fälle, in denen (gewaltsames) Sterben beschrieben wird, darauf hin untersucht werden, wie sich die Anteile der Ausdrucksmöglichkeiten von *STERBEN* im Untersuchungszeitraum verändern. Wenn dann immer noch eine Zunahme des Ausdrucks *in den Tod* (*reißen*) festgestellt wird, kann vorläufig These 2 angenommen werden.

Im skizzierten Fall definiere ich ein Teilkorpus  $K'$ , das aus Artikeln des Ausland-Ressorts besteht, in denen mindestens einer der folgenden Ausdrücke vorkommt: *Opfer*, *ums Leben*, *(ge)tötet*, *sterben* (in allen Flexionsformen) und *in den Tod*. Dieses Teilkorpus enthält damit potenzielle Verwendungskontexte von *in den Tod*. Im Anschluss werden die paradigmatischen Konstellationen in diesem Teilkorpus

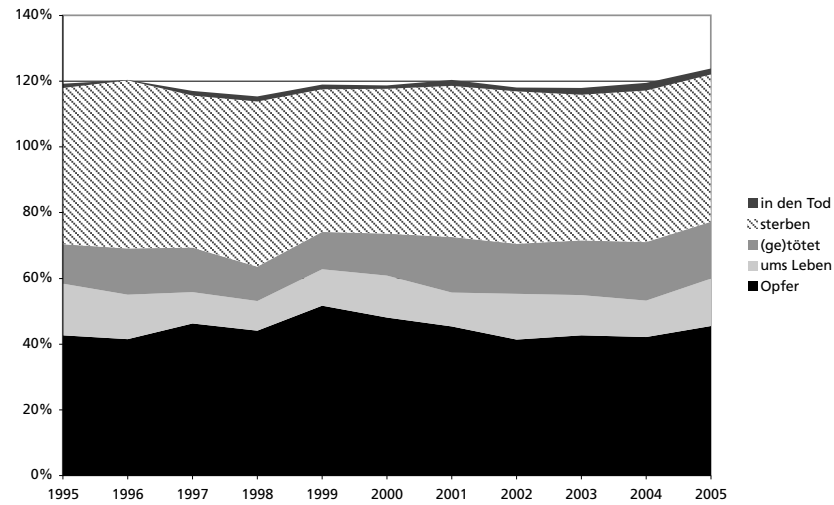


**Abbildung 8.6:** Paradigmatische Konstellationen von STERBEN im Ressort ‚Ausland‘: Die Verteilung der Ausdrucksvarianten zu STERBEN in Artikeln desselben Kontextes im Ressort ‚Ausland‘. Pro Artikel können mehrere Ausdrucksvarianten vorkommen.

geschieden nach Jahrgängen beschrieben und man erhält so die Zusammenstellung, wie sie Abbildung 8.6 darstellt. Die Grafik zeigt die Verteilung der Ausdrucksvarianten in Relation zum Teilkorpus. Da in einem Artikel mehrere Ausdrucksvarianten vorkommen können, ergeben sich pro Jahrgang insgesamt Werte über 100%.

Im Vergleich dazu zeigt Abbildung 8.7 auf der nächsten Seite die paradigmatischen Konstellationen von STERBEN in allen Ressorts der Zeitung. Im Kontrast zu Abbildung 8.6 wird deutlich, dass auch in Relation zu potenziellen Verwendungskontexten im Ressort ‚Ausland‘ nach dem Jahr 2001 vermehrt *in den Tod* verwendet wurde. Diese Zunahme ist aber nur für dieses Ressort typisch. In den anderen Ressorts gibt es über die ganze Zeitperiode hinweg gesehen Verwendungen dieses Ausdrucks.

Die beiden Grafiken zeigen aber auch, dass im Auslands-Ressort die Verwendung von Ausdrücken des Paradigmas STERBEN leicht zugenommen haben (vgl. dazu auch Tabelle 8.5 auf Seite 167), während über alle Ressorts hinweg gesehen die Verwendung etwa gleich blieb.



**Abbildung 8.7:** Paradigmatische Konstellationen von STERBEN über alle Ressorts: Die Verteilung der Ausdrucksvarianten zu STERBEN in Artikeln desselben Kontextes. Pro Artikel können mehrere Ausdrucksvarianten vorkommen.

In Bezug auf die oben formulierten Thesen kann Folgendes festgehalten werden:

- Beobachtung 1: Über alle Ressorts hinweg haben Artikel, in denen eine Ausdrucksvariante des Paradigmas STERBEN verwendet wird, nicht zugenommen. Anders im Ressort ‚Ausland‘: Da hat die Verwendung seit 1998, und vor allem mit dem Jahr 2001, zugenommen. Damit kann mit gutem Grund vermutet werden, dass die potenziellen Verwendungskontexte sich verschoben haben: Sie sind vermehrt in Ausland-Themen zu finden.
- Beobachtung 2: Im Ressort ‚Ausland‘ nahmen die Ausdrucksvarianten *(ge)tötet* und *in den Tod* seit 2001 zu. Das ist das Spiegelbild der vermehrten Thematisierung von STERBEN in diesem Ressort.
- Fazit: Die Beobachtungen geben Hinweise darauf, dass im internationalen, politischen Kontext vermehrt von tödlichen Ereignissen die Rede ist. Zudem hat sich die Art und Weise, wie

Jahr	STERBEN		Total
	#	%	
1995	96	19,35	496
1996	110	22,31	493
1997	100	20,08	498
1998	83	16,47	504
1999	106	18,76	565
2000	116	19,59	592
2001	134	23,51	570
2002	120	23,58	509
2003	145	25,80	562
2004	122	24,02	508
2005	117	24,07	486
Total	1249		5783
$\chi^2 = 26$ , $df = 10$			
$p < 0,01$ (signifikant)			

**Tabelle 8.5:**  $\chi^2$ -Statistik zur Verwendung von Ausdrücken des Paradigmas STERBEN in der Auslandsberichterstattung des NZZ-Korpus (Einheit: Artikel). Zum Paradigma gehören folgende Ausdrücke: *Opfer*, *ums Leben*, *(ge)tötet*, *sterben* (in allen Flexionsformen) und *in den Tod*.

darüber gesprochen wird, leicht verändert: Die Ausdrucksvariante *in den Tod* (*gerissen*) bzw. (*riss*) *in den Tod* hat signifikant zugenommen.

Es ist naheliegend, Ereignisse wie den Irak-Krieg, den Konflikt im Nahen Osten oder die Attentate der Kaiba als Grund für die Zunahme von *in den Tod*-Formulierungen im Auslands-Ressort anzusehen. Dies zeigt z. B. folgender Beleg:

- (12) Die Explosion riss 47 Personen **in den Tod** und brachte über 100 anderen zum Teil schwere Verletzungen bei. *Neue Zürcher Zeitung vom 15. September 2004, Ressort ‚Ausland‘: „Verheerender Anschlag auf Bagdader Markt Zahlreiche Tote – Urhebererklärung der Bande Zarkawis“.*

Eine Analyse des Kontextualisierungsprofils (vgl. Kapitel 8.2.2) von *in den Tod* müsste nun weiter zeigen, in welchen Kontexten dieser Ausdruck normalerweise verwendet wird.

Paradigmatische Konstellationen sind ein besonders geeigneter Schlüssel zur Analyse des Sprachgebrauchs, denn die Spezifität der

Sprechweise in einem Diskurs liegt oft in der Selektion *einer* möglichen Variante aus einem semantischen Paradigma.

#### 8.2.4 Pragmatische Funktionen

Durch die Kontrastierung von Kollokationen nach Teilkorpora gewinnt man oft Muster, die typisch für ein Teilkorpus sind (vgl. Kapitel 8.1). Scheiden sich die Teilkorpora beispielsweise synchron an Textsorten, können mit den gefundenen Mustern bestimmte pragmatische Funktionen in Verbindung gebracht und z. B. die folgende These formuliert werden: Wenn das Muster X in der bestimmten Textsorte S auftritt, ist damit sehr wahrscheinlich eine bestimmte pragmatische Funktion f verbunden. Tritt das Muster X hingegen in einer anderen Textsorte auf, ist dies nicht der Fall.

Am Beispiel von Leserbriefen lässt sich das zeigen (Bubenhof 2008b): Eine Reihe von im untersuchten Korpus<sup>11</sup> berechneten Kollokationen scheint im Kontext von Leserbriefen sehr differenzierte pragmatische Funktionen aufzuweisen. So bedeutet *ein Dorn im Auge* dort nicht bloß „jmdm. ein Ärgernis, unerträglich sein“, wie die Duden-Definition lautet (Duden 2002, ‚Dorn‘), sondern die Wendung unterstellt oft einem Gegner eine unlautere Argumentation in einer Sache. Oder anders ausgedrückt: Der Verfasserin/der Verfasser des Leserbriefs unterstellt der Gegnerschaft fehlende Wahrhaftigkeit (Eggler 2006, 45f.) in der Argumentation. Es handelt sich um einen Aufruf, die Argumentation der Gegnerschaft nicht ernst zu nehmen, da, so die Unterstellung, die Argumentation nur eine pauschale (und damit ungerechtfertigte) Abneigung gegenüber der Sache verschleierte. Das folgende Beispiel aus Bubenhof (2008b) mag das illustrieren:

- (13) Seit langem sind den rechtsgerichteten Kreisen das Verbandsbeschwerderecht sowie die Lex Koller (Limitierung von Wohneigentum durch Ausländer) **ein Dorn im Auge**. Man scheut sich dabei nicht, mit Unwahrheiten die breite Masse zu manipulieren. Nur gerade ein Prozent aller Verwaltungsgerichtsbeschwerden stammt von Umweltschutzorganisationen. Die

<sup>11</sup> Das Korpus umfasst alle Leserbriefe, die zwischen dem 1. Januar und dem 31. August 2006 im Zürcher ‚Tages-Anzeiger‘ erschienen sind. Es handelt sich um 318 535 Wörter (1357 Leserbriefe). Die Kollokationen wurden corpus-driven aus dem Korpus berechnet (Bubenhof 2008b).

Behauptung, die Verbände würden die Bautätigkeit behindern, ist damit klar widerlegt. *Tages-Anzeiger vom 1. März 2006, Ressort ‚Leserbriefe‘: „Hickhack um Beschwerde: Bautätigkeit wird nicht verhindert“.*

Der Verfasser oder die Verfasserin nimmt in dieser Passage ein Argument der *rechtsgerichteten Kreise* gegen das Verbandsbeschwerderecht auf, das behauptet, Verbände würden die Bautätigkeit behindern. Damit wird das Argument mit der Behauptung widerlegt, nur gerade ein Prozent aller Beschwerden würden von Verbänden stammen. Für den Verfasser/die Verfasserin ist somit klar, dass es keine vernünftigen Argumente gegen das Verbandsbeschwerderecht gibt. Trotzdem kämpften die rechtsgerichteten Kreise dagegen an, weil das Recht ihnen *ein Dorn im Auge* sei. Damit erhält dieser Ausdruck in diesem Kontext die Bedeutung ‚Dagegen sein, ohne ein vernünftiges Argument anzugeben‘.

Folgerichtig würde man – so die These – in Leserbriefen nur sehr selten die Verwendung von *ein Dorn im Auge* in Verbindung mit einem Pronomen in der ersten Person Sg. finden: *Mir ist X ein Dorn im Auge*. Das würde in Leserbriefen wenig Sinn machen, da die Textsorte ein Argumentieren erfordert. Deswegen wird mit dieser Wendung meistens einem Gegner unterstellt, ihm sei etwas *ein Dorn im Auge*, wobei damit die Argumentation des Gegners als unlauter diffamiert wird.

Das Beispiel zeigt, wie eine pragmatische Funktion (ARGUMENTIEREN, DIFFAMIEREN) als Art der Distribution eines Musters in bestimmten Korpora beschrieben werden kann. Somit steht auch eine Klassifizierung von Mustern nach pragmatischen Kriterien zur Verfügung, die sich korpuslinguistisch operationalisieren lässt: Stimmt die erwartete Distribution auch mit der beobachteten in anderen Korpora überein, bewährt sich die Klassifizierung.

Auch Tognini-Bonelli (2001, 29ff.) zeigt, wie pragmatische Funktionen von bestimmten Wörtern korpuslinguistisch herausgearbeitet werden können. Sie zeigt dies nicht am Beispiel einer Kollokation, sondern an einem Einzelwort: *except*. Auch sie betont, dass bestehende Grammatiken kaum Auskunft über die kommunikative Funktion von Wörtern geben. Die Korpusanalyse zeigt aber, dass die pragmatische Funktion von *except* sehr wohl differenziert erklärt werden kann.



Denn es werden zwei gegensätzliche Funktionen (Belege 14 und 15) sichtbar:

- (14) Pluto also has its own moon, Charon, of which hardly anything is known **except that** it is anomalously large, weighing a sixth as much as Pluto itself. *Zit. nach Tognini-Bonelli (2001, 31).*

Die generelle Behauptung ‚kaum etwas von Charon ist bekannt‘ wird durch die Ausnahme ‚wir wissen nur eines: er ist ungewöhnlich groß‘ nicht ungültig. Sondern die Ausnahme wird in den Wahrheitswert der ersten Behauptung integriert.

- (15) ... say that to trust Mr Hussein is to put their heads in a crocodile's jaws. They are right – **except that** the crocodile in question is largely defanged, with its tail locked in a trap. *Zit. nach Tognini-Bonelli (2001, 31).*

Hier wird mit der Formulierung der Ausnahme (‚das Krokodil ist abgeschwächt und sein Schwanz in einer Falle gefangen‘) die metaphorische Gültigkeit der Behauptung umgekehrt: Es ist also *nicht* gefährlich, Hussein zu trauen. Die Ausnahme macht die Behauptung unwahr.

„These are two examples that seem to make use of the same grammatical structure in order to convey a communicatively opposite meaning.“ (Tognini-Bonelli 2001, 31) Beide Male wird dieselbe Formulierung *except that* verwendet, allerdings mit unterschiedlichen pragmatischen Funktionen, die über die Semantik von *except* hinaus gehen. Im zweiten Fall ist *except that* die Schlüsselformulierung in einer Argumentation. Im ersten Fall wird die Formulierung nicht in einem argumentativen Kontext verwendet.

Die beiden Gebrauchsvarianten bestätigen sich bei der weiteren Analyse des Korpus. Und es zeigt sich auch: Die erste, nicht-argumentative Variante, ist weit weniger häufig anzutreffen als die argumentative Variante. Trotzdem wird diese in Grammatiken weitgehend ignoriert.<sup>12</sup>

<sup>12</sup> Vgl. auch argumentative Funktion von *broadly* im Gegensatz zu *largely* in Tognini-Bonelli (2001, 39).

Die pragmatischen Funktionen von solchen Gebrauchsvarianten könnten sprachspezifisch sein. So zeigt sich im NZZ-Korpus bei der Suche nach *ausser(,) dass* ein anderes Bild: Alle 19 Treffer entsprechen dem ersten Fall, bei dem die Ausnahme in den Wahrheitswert der generellen Behauptung integriert wird. Auch eine weitere, unsystematische Recherche in deutschsprachigen Zeitungen bringt keine Belege für *ausser/außer(,) dass* im Sinne des zweiten Falls hervor, außer einem, der allenfalls dem zweiten Typus zugeordnet werden könnte:

- (16) Alles in Ordnung. Alles normal. Ausser, dass man sich zehntausend Meter über dem Atlantik befindet [...]. Ausser, dass man mit wildfremden Menschen in einem künstlich unter Druck gehaltenen Raum die Nacht verbringt. [...] Ausser, dass das Ding fliegt. Nichts ist in Ordnung, und nichts ist normal. [...] Das Verkehrsmittel Flugzeug existiert nur dank einer weltweiten Verschwörung zur Vorspiegelung von Normalität während einer absoluten Extremsituation. Ohne sie würde kein normaler Mensch jemals ein Flugzeug betreten. *NZZ Folio 06/96, Thema: Vom Reisen. „Fly away: Über die Normalität des Anomalen“ von Martin Suter, Schriftsteller.*

Es ist wahrscheinlich symptomatisch, dass es sich hierbei um einen literarischen Text handelt, in dem die zweite Gebrauchsvariante erscheint. Das von Tognini-Bonelli (2001, 29ff.) beschriebene Phänomen scheint (zumindest in Zeitungstexten) im Deutschen nicht in großem Maße vorzukommen.

### 8.3 Überprüfen: Zurück in die Daten

Im Kapitel 8.2 wurde dargelegt, welche Kategorien und Ansätze für die Interpretation der Daten vorgeschlagen werden. Dabei zeigte sich, dass der Schritt der Interpretation bereits eine Scharnierfunktion zwischen dem corpus-driven- und dem corpus-based-Paradigma einnimmt, da die Interpretation mit einem erneuten – und diesmal klar gesteuerten – Blick auf die Korpusdaten einhergeht.

Ziel des interpretativen Schritts war, eine Diskursbeschreibung aus den Beobachtungen zu abstrahieren. Um die Adäquatheit der Beschreibung zu überprüfen, sind die beiden folgenden Schritte möglich:

1. Überprüfung der Diskursbeschreibung anhand des Untersuchungskorpus: Haben die gefassten Thesen auch bei erneuter Prüfung Bestand? Bewähren sich die corpus-driven gemachten Erklärungen auch aus corpus-based-Sicht oder muss das Bild ggf. korrigiert werden?
2. Überprüfung der Diskursbeschreibung anhand eines Vergleichskorpus: Je nach Untersuchungsinteresse kann das Vergleichskorpus so gewählt werden, dass es bezüglich Kriterien wie Thema, Textsorte oder sprachlicher Varietät dem Untersuchungskorpus möglichst ähnlich ist, oder eben gerade im Kontrast dazu steht.

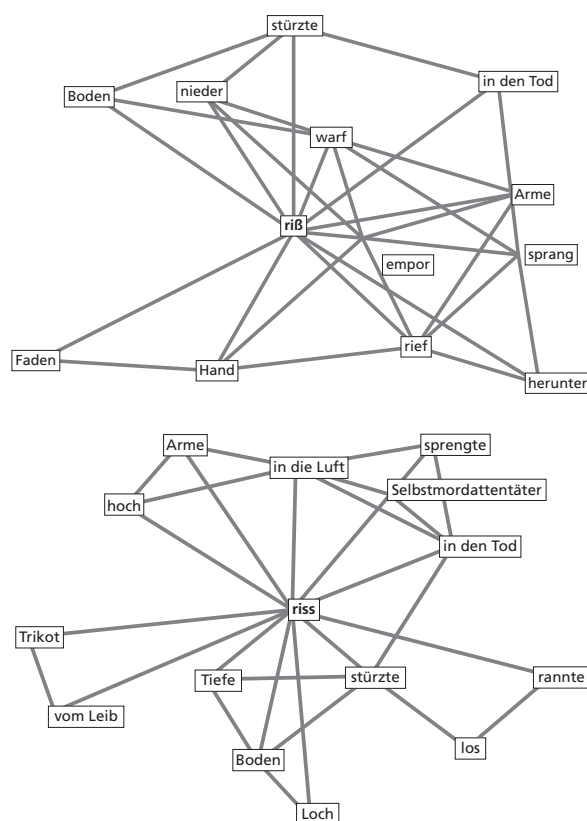
Die auf Seite 164 formulierte These, dass die Zunahme von *in den Tod* in der NZZ im Jahr 2005 auf eine Veränderung der paradigmatischen Konstellation von STERBEN zurück geht, kann anhand eines großen Vergleichskorpus<sup>13</sup> zumindest ansatzweise überprüft werden.

Als Beispiel sei an dieser Stelle bloß auf die Abbildung 8.8 auf der nächsten Seite verwiesen. Sie zeigt die im Wortschatz-Korpus (Wortschatz Leipzig o.J.) berechneten signifikanten Kollokatoren zu *riß* bzw. *riss*. Bei diesem Beispiel erweist sich die ab dem Jahr 2000 schrittweise eingeführte neue deutsche Rechtschreibung als hilfreich, die das Korpus bezüglich dieses Wortes temporal grob differenzierbar macht. Das Wort *riß/riss* wird nach der Darstellung in beiden Schreibvarianten in gemeinsamen (*in den Tod*) aber auch unterschiedlichen Kontexten verwendet. Neue Schreibung (ca. ab 2000) (*riss*): *sprengte, in die Luft* und *Selbstmordattentäter*. Alte Schreibung (ca. vor 2000) (*riß*): *an sich, Faden, Tiefe* etc.

Die auf Seite 164 gemachte Beobachtung einer Zunahme der Wendung *in den Tod* (*reißen*) in der NZZ des Jahres 2005 gegenüber den Jahren 1995–2004 im Kontext von Bomben- und Selbstmordattentaten spiegelt sich also auch im viel größeren Wortschatz-Korpus wieder.

Grundsätzlich problematisch bei Vergleichen dieser Art sind oft die unterschiedlichen Strukturen von eigenem und Vergleichskorpus, z. B.

<sup>13</sup> Zu den verfügbaren Ressourcen vgl. Kapitel 9. Zu denken ist z. B. an das Deutsche Referenzkorpus des IDS (DeReKo IDS o.J.), das Digitale Wörterbuch der deutschen Sprache des 20. Jahrhunderts (DWDS o.J.) oder das Internet-Korpus des Wortschatzprojekts der Universität Leipzig (Wortschatz Leipzig o.J.). Daneben kann auch das Web als Vergleichskorpus dienen. Zu den Möglichkeiten und Problemen damit vgl.: Baroni/Bernardini (2006), Bickel (2006), Jung (2005), Kilgarriff/Grefenstette (2003), Mautner (2005), Sharoff (2006), Thelwall (2005), Volk (2002).



**Abbildung 8.8:** Signifikante Kollokatoren zu *riß* bzw. *riss* im Wortschatz Leipzig (o. J.)-Korpus, berechnet am 21. April 2007. Grafik NB auf Basis der Originalgrafik.

bezüglich Textsorte, Systematik der Metainformationen zu den Texten etc., und/oder die fehlende Transparenz diesbezüglich im Vergleichskorpus. Oder es ist nicht möglich, die Abfragen und Berechnungen im Vergleichskorpus derart genau zu beeinflussen, um methodisch der Analyse des eigenen Korpus möglichst zu entsprechen. In der Praxis müsste für Vergleichsuntersuchungen im Referenzkorpus oft eine Teilmenge nach Kriterien wie Publikationsdatum oder Textsorte definiert werden, um diesbezüglich vergleichende Aussagen machen zu können. Im Deutschen Referenzkorpus des IDS (DeReKo IDS o. J.) ist das gut

möglich, im DWDS (o.J.) nur bedingt und im Wortschatz-Korpus (Wortschatz Leipzig o.J.) überhaupt nicht. Sinnvoll sind Vergleiche der Art ‚Spezialkorpus‘  $\leftrightarrow$  ‚allgemeiner Sprachgebrauch in einem Referenzkorpus‘, die die Frage verfolgen, ob ein beobachtetes Phänomen im eigenen Korpus A auf einen allgemeineren Sprachgebrauch (repräsentiert durch ein Referenzkorpus B) übertragen werden kann.

## 9 Ressourcen der Korpus- und Computerlinguistik

Die folgende Übersicht über Ressourcen der Korpus- und Computerlinguistik kann nur den momentanen Stand widerspiegeln, der sich in gewissen Bereichen rasch ändern wird. Trotzdem sei hier auf eine Reihe von Korpora und Software verwiesen, die entweder im Zusammenhang mit der vorliegenden Arbeit verwendet wurden oder aber sich ebenfalls gut für ähnlich gelagerte Untersuchungen eignen würden.

### 9.1 Öffentlich verfügbare Referenzkorpora

Die bisherigen Ausführungen haben klar gemacht, dass die Zusammenstellung eines eigenen Korpus meistens unumgänglich ist. Kapitel 8.3 machte jedoch auch deutlich, dass bestehende Korpora für die Überprüfung des Modells wichtig sind. Nämlich im Sinne von Referenzkorpora, die relativ unspezifisch versuchen „deutsche Gegenwartssprache“ (DeReKo IDS o.J.) abzudecken und sich so von domänenspezifischen eigenen Korpora unterscheiden.

Tabelle 9.1 auf Seite 178 zeigt die wichtigsten öffentlich verfügbaren deutschsprachigen Korpora geschriebener Sprache im Vergleich. Daneben existieren eine Vielzahl kleinerer oder spezialisierter Korpora, teilweise mit beschränkten Zugangsmöglichkeiten. Es sei hier für eine umfassendere Übersicht auf Lemnitzer/Zinsmeister (2006, 107ff.) sowie auf einschlägige Web-Portale verwiesen: Breyer (2005), Bubenhofer (2006).

Die Übersicht in Tabelle 9.1 auf Seite 178 zeigt, wo die Stärken und Schwächen der drei Korpora liegen. Was die Kontrollmöglichkeiten der Abfragen betrifft, erlaubt das DeReKo IDS (o.J.)-Korpus über die Schnittstelle COSMAS II am meisten Optionen, gefolgt vom DWDS (o.J.)-Korpus. Das Wortschatz Leipzig (o.J.)-Projekt hingegen erlaubt das Herunterladen des gesamten Korpus, was wiederum eigene Analysen ermöglicht. Dies ist allerdings mit dem Nachteil ver-

bunden, dass aus lizenzrechtlichen Gründen in diesem Korpus die Texte nicht integral vorhanden sind, sondern in einzelne Sätze aufgelöst und so nicht mehr als zusammengehörend identifizierbar sind. Im DWDS (o. J.)-Korpus sind diachrone Analysen besonders einfach zu realisieren, da das Korpus über die Dekaden eine gleichmäßige Textsortenschichtung aufweist.

Das DWDS (o. J.)-Korpus bietet eine Annotation der Wortarten. Damit ist es möglich, Suchanfragen wie *Kampf mit ... [Substantiv]* zu formulieren. Aufgrund der Überlegungen zu Lemmatisierung und Wortartenbestimmung ist jedoch fraglich, wie wichtig die Annotation ist (vgl. Kapitel 6.5). Mängel zeigen sich beim DWDS (o. J.)-Korpus bei der Berechnung von Kollokationen, die nur auf der Basis von Lemmata, nicht aber von Wortformen oder Wortgruppen gemacht werden können.

Neben den oben erwähnten drei Korpora, die extra für linguistische Bedürfnisse erstellt worden sind, gibt es weitere Datenquellen, die für eine korpuslinguistische Diskursanalyse von Interesse sind. Dazu gehört das WWW, worin über Suchmaschinen wie ‚Google‘ recherchiert werden kann. Natürlich ergeben sich einige Schwierigkeiten bei der Nutzung des WWW als Datenquelle. Zu den größten Hürden gehören die fehlenden Informationen zum Umfang des von den Suchmaschinen erfassten Datenbestandes (in Relation zur Grundgesamtheit aller Webseiten), die heterogene Menge an Textsorten und die limitierten Suchmöglichkeiten über die Suchmaschinen. Trotzdem erweist sich das WWW punktuell als nützliche Datenquelle.<sup>1</sup>

Eine weitere wichtige Alternative zu linguistischen Korpora sind Zeitungsdatenbanken, die z. B. über Universitätsbibliotheken zugänglich sind. Meist seit den 1990er-Jahren sind viele Zeitungen und Zeitschriften in elektronischer Form erhältlich. Die Recherchemöglichkeiten erlauben normalerweise die Volltextsuche, sowie die Suche nach Publikationsdatum, Ressort, Autor- und Titelinformationen etc.

In der vorliegenden Arbeit werde ich corpus-driven generierte Hypothesen hauptsächlich mit Recherchen im DeReKo IDS (o. J.)-Korpus, im WWW und in der Zeitungsdatenbank LexisNexis (o. J.)

<sup>1</sup> Vgl. zur Evaluation des WWW als Korpus Baroni/Bernardini (2006), Bickel (2006), Kilgarriff/Grefenstette (2003), Volk (2002), speziell im Bereich der Diskursanalyse Jung (2005), Mautner (2005) und zu technisch-methodischen Hinweisen Fletcher (2007), Sharoff (2006), Thelwall (2005).

überprüfen. Die Zusammensetzung des DeReKo IDS (o.J.)-Korpus ist sehr transparent und kann durch die Bildung von ‚virtuellen Korpora‘ beeinflusst werden. Zudem besteht das Korpus, vor allem für die Periode ab 1990, zu großen Teilen aus Zeitungstexten, was die Vergleichbarkeit mit dem NZZ-Korpus steigert. Die Zeitungsdatenbank LexisNexis (o.J.) enthält die wichtigsten deutschsprachigen Tages- und Wochenzeitungen ab den 1990er-Jahren im Volltext und ermöglicht eine rasche Recherche nach Belegen für bestimmte Phänomene.



<i>Name</i>	<b>Das Deutsche Referenzkorpus (DeReKo)</b>	<b>Das Digitale Wörterbuch der deutschen Sprache des 20. Jh. (DWDS)</b>	<b>Wortschatz Universität Leipzig</b>
<i>Umfang (Token)</i>	2 Mrd. intern, 940 Mio. öffentlich	100 Mio.	30 Mio. Sätze <sup>2</sup>
<i>Downloadbar</i>	nein	nein	ja
<i>Textsorten</i>	vor allem ab 1990 Schwergewicht auf Zeitungstexte, vorher auch Belletristik; Gebrauchstexte	pro Dekade jeweils ca. 26 % Schöne Literatur, 27 % journalistische Prosa, 22 % Fachprosa, 20 % Gebrauchstexte, 5 % Texte gesprochener Sprache	Zeitungstexte (online)
<i>Zeitraum</i>	hauptsächlich 1980–heute; teilweise auch ältere Texte	1900–2000	unbekannt
<i>Teilkorpora</i>	nach beliebigen Kriterien definierbar	nach Jahr und Textsorte filterbar	nein
<i>max. Beleggröße</i>	Absatz	mehrere Sätze	Satz
<i>Metainformationen</i>	Datum, Quelle	Datum, Quelle	keine
<i>Annotation</i>	kleiner Teil mit Wortarten	Wortarten vollständig	nein
<i>Erschließung</i>	COSMAS II (Web- oder Windows-Client): komplexe Abfragen möglich	DDC als Web-Client: komplexe Abfragen möglich	Web-Client: Lemma-Suche

Tabelle 9.1: (Fortsetzung folgende Seite)

<sup>2</sup> Diese Angabe wird in Quasthoff u. a. (2006, 1800) gemacht; Duffner/Näf (2006, 9) sprechen von 500 Mio. Token.

<i>Name</i>	<b>Das Deutsche Referenzkorpus (DeReKo)</b>	<b>Das Digitale Wörterbuch der deutschen Sprache des 20. Jh. (DWDS)</b>	<b>Wortschatz Universität Leipzig</b>
<i>Kollokationen</i>	beliebig berechenbar; daneben möglich: Nutzung der Kookkurrenzdatenbank CCDB (Belica 2001–2006)	berechenbar nach fünf Methoden	berechenbar auf Basis Lemma
<i>Phrasensuche</i>	ja	ja	nein
<i>Zugang</i>	<a href="http://www.ids-mannheim.de/">http://www.ids-mannheim.de/</a>	<a href="http://www.dwds.de">http://www.dwds.de</a>	<a href="http://corpora.informatik.uni-leipzig.de/">http://corpora.informatik.uni-leipzig.de/</a>

**Tabelle 9.1:** Die wichtigsten öffentlich verfügbaren deutschsprachigen Korpora im Vergleich. Gemäß den Angaben der Korpora (DeReKo IDS o. J., DWDS o. J., Wortschatz Leipzig o. J.) sowie Duffner/Näf (2006) und Quasthoff u. a. (2006).

## 9.2 Software zur Korpusabfrage

In den im vorherigen Kapitel erwähnten (linguistischen) Korpora sind gleich auch sog. ‚Corpus Query Tools‘ enthalten, mit denen die Korpora nach bestimmten Kriterien abgefragt werden können. Für das DeReKo IDS (o.J.)-Korpus steht ‚COSMAS II‘, für das DWDS (o.J.) ‚DDC‘ zur Verfügung. Die beiden Query-Tools erlauben komplexe Abfragen, für die eine eigene Sprache zur Formulierung der Suche zur Verfügung steht. Ich kann hier nicht im Detail auf diese Sprachen eingehen und verweise stattdessen auf die entsprechenden Dokumentationen der beiden Korpora, sowie auf Bubenhofer (2006, ‚COSMAS II‘).

Für eigene Korpora existieren einige Query-Tools, auch ‚Concordancer‘ genannt, da sie vornehmlich nach der Eingabe eines Suchausdrucks die Treffer in Form von sog. ‚Konkordanzen‘ darstellen: Sie zeigen also alle Kontexte des Suchausdrucks an den verschiedenen Stellen des Korpus auf.<sup>3</sup> In einschlägiger Literatur finden sich zahlreiche Zusammenstellungen solcher Software (Bubenhofer 2006, Lemnitzer/Zinsmeister 2006, Scherer 2006).

Für das in der vorliegenden Arbeit benutzte NZZ-Korpus habe ich ein eigenes Datenbank- und Abfragesystem entwickelt. Die Korpus-Datenbank<sup>4</sup> ist über ein PHP-Interface<sup>5</sup> abruf- und steuerbar. Es erfüllt folgende Funktionen:

- Durchsuchen der Datenbank mit der SQL-Abfragesprache.
- Sortierung der Ergebnisse nach unterschiedlichen Kriterien.
- Gruppierung der Ergebnisse nach Kategorien wie ‚Jahr‘, ‚Resort‘ etc.
- Ausgabe der Ergebnisse als KWiC-Zeilen (‚Key Word in Context‘), Titelzeilen oder Volltexte.

<sup>3</sup> Vgl. in diesem Zusammenhang auch die Bezeichnung ‚KWiC‘ (‚Key Word in Context‘). Damit ist eine Darstellung gemeint, bei der der Suchausdruck pro Treffer jeweils in der Mitte einer Zeile angezeigt wird und links und rechts der unmittelbare Kontext in einer bestimmten Länge anschließt.

<sup>4</sup> Es handelt sich um eine MySQL-Datenbank.

<sup>5</sup> Das PHP-Interface zur Verwaltung des NZZ-Korpus kann beim Autor bezogen werden: <http://www.bubenhofer.com/korpusanalyse/>.

- Statistische Berechnungen zu den Ergebnissen: Lage- und Streuungsmaße (vgl. Kapitel 7.1), aber auch Berechnung der Signifikanz von Verteilungen (vgl. Kapitel 7.2).

Alternativ zu dieser Datenbanklösung wäre es möglich, die Daten in einem XML-Format, z. B. nach TEI (‘Text Encoding Initiative’)<sup>6</sup>, zu codieren und eine Konkordanz-Software wie ‚Xaira’<sup>7</sup> zur Abfrage zu verwenden. Diese Lösung würde es ebenfalls ermöglichen, die Metadaten zu den Artikeln in die Korpusabfrage einzubeziehen. Allerdings ergaben sich bei meinen Testläufen damit Geschwindigkeitsprobleme, sodass ich darauf verzichtete. Zudem läge der Vorteil der XML-Codierung primär in der Transparenz des Formats, das eine Weitergabe der Daten für andere Nutzungen vereinfachte. Da eine Weitergabe der Daten aus lizenzrechtlichen Gründen sowieso nicht möglich ist, wird dieser Vorteil hinfällig.

Als modular aufgebautes und leistungsfähiges Softwarepaket wäre auch die ‚IMS Corpus Workbench (CWB)’ (Christ/Schulze 1995) als Corpus Query Tool in Frage gekommen. Es wird auch als besonders geeignet für data-driven-Analysen angepriesen und wäre deshalb meinen methodischen Interessen entgegen gekommen. Allerdings war es in meinem Fall einfacher, eine eigene, maßgeschneiderte Lösung umzusetzen, die exakt die benötigten Funktionen bietet.

### 9.3 Software für die Berechnung von Kollokationen

Die Berechnung von Kollokationen oder Mehrworteinheiten ist im Prinzip eine simple Zählaufgabe, die besonders gut maschinell erledigt werden kann. Auch programmiertechnisch ist das Problem eigentlich trivial und ließe sich sogar mit Standard-Unixbefehlen erledigen:

```
tail +2 wordlist > nextwords
paste wordlist nextwords > bigrams
sort -d bigrams | uniq -c > bigrams_sorted
tail +3 wordlist > nextnextwords
paste wordlist nextwords nextnextwords > trigrams
sort -d trigrams | uniq -c > trigrams_sorted
...
```

<sup>6</sup> Vgl. <http://www.tei-c.org/> (14. 2. 2008).

<sup>7</sup> Vgl. <http://www.oucs.ox.ac.uk/rts/xaira/> (14. 2. 2008).

Dabei ist `wordlist` das Korpus in Wortlistenform, so dass pro Zeile bloß ein Wort steht. Auch mit einer Programmiersprache wie Perl lässt sich schnell ein entsprechendes Script erstellen, das alle möglichen Wortgruppen sammelt, auflistet, sortiert und zählt.

Aber man kann auch auf bestehende Programme zurückgreifen, die größeren Komfort bieten. In Tabelle 9.2 auf der nächsten Seite sind drei Programme aufgeführt, die sich besonders für eine corpus-driven-Berechnung von Mehrworteinheiten eignen. Für die vorliegende Arbeit verwendete ich, neben eigenen Scripts, das ‚Ngram Statistics Package‘. Daneben experimentiere ich aber auch mit ‚kfNgram‘ und ‚Concgram‘.

Die aufgeführten Programme sind nicht die einzig existierenden. Oft bieten auch Konkordanz-Programme Kollokationsberechnungen an, meistens jedoch nur corpus-based, was bedeutet, dass das Programm von einem bestimmten Ausdruck ausgeht, zu dem dann die Kollokatoren berechnet werden. Ausnahmen bilden da das kommerzielle Windows-Programm ‚Wordsmith‘ und das kostenlose ‚AntConc‘, das für verschiedene Betriebssysteme zur Verfügung steht.<sup>8</sup>

---

<sup>8</sup> Detailliertere und aktuelle Informationen zu diesen Programmen in Bubenhofer (2006).

<i>Name</i>	<b>Ngram Statistics Package (NSP)</b>	<b>kfNgram</b>	<b>Concgram</b>
<i>Autoren</i>	Ted Pedersen	William H. Fletcher	Chris Greaves
<i>diskontinuierliche Wortgruppen</i>	ja	ja, über ‚Phrase-Frames‘ <sup>9</sup>	ja
<i>freie Reihenfolge der Wortgruppen</i>	nein	nein	ja
<i>statistische Maße</i>	Bibliothek an 13 Maßen für Bigramme und 4 für Trigramme; Bibliothek erweiterbar	nur Frequenz	ja: t-score und Mutual Information
<i>Besonderes</i>	Perl-Programme, dadurch sehr flexibel in eigene Abläufe integrierbar	Geschickte Algorithmen und effiziente Programmiersprache machen kfNgram sehr schnell	Komfortable Lösung zur Berechnung der freiesten Form von Wortgruppen: Concgrams. Da sehr rechenintensiv bei großen Korpora langsam.
<i>Betriebssysteme</i>	Mac OS X, Unix-Systeme, Windows	Windows	Windows
<i>Lizenz/Kosten</i>	GNU General Public License (GPL), kostenlos	kostenlos	US \$ 15
<i>Bezug</i> (Stand: 29. 2. 2008)	<a href="http://www.d.umn.edu/~tpederse/nsp.html">http://www.d.umn.edu/~tpederse/nsp.html</a>	<a href="http://www.kwicfinder.com/kfNgram/kfNgramHelp.html">http://www.kwicfinder.com/kfNgram/kfNgramHelp.html</a>	<a href="http://www.edict.com.hk/pub/concgram/">http://www.edict.com.hk/pub/concgram/</a>
<i>Literatur</i>	Banerjee/Pedersen (2003)		Cheng u. a. (2006)

**Tabelle 9.2:** Eine Auswahl an Software zur corpus-driven Berechnung von Kollokationen.

<sup>9</sup> Die Berechnung diskontinuierlicher Wortgruppen erfolgt über die Berechnung von sog. ‚Phrase-Frames‘: Im Anschluss der Berechnung kontinuierlicher n-Gramme werden Wortgruppen gesucht, die jeweils bis auf eine oder mehrere Token identisch sind.



### **III**

## **Anwendungsbeispiele**





In Teil II der vorliegenden Arbeit habe ich die Methoden einer korpuslinguistischen Diskursanalyse vorgeschlagen, die aufgrund der Beobachtung von Sprachgebrauchsmustern im Wechsel zwischen corpus-driven und corpus-based versucht, Diskurse zu beschreiben. Im Kern ist diese Methode in Kapitel 5.2 dargestellt.

Im letzten Teil der vorliegenden Arbeit möchte ich nun nicht eine abgeschlossene Diskursanalyse zum Thema X leisten, sondern anhand einiger Beispiele aufzeigen, wie die vorgeschlagene Methode für unterschiedliche Untersuchungsinteressen eingesetzt werden kann. Ich werde allerdings nur einen Bruchteil der gewonnenen Daten zu Mustern abstrahieren und deren Potenzial für diskursanalytische Fragestellungen aufzeigen. Dass damit dieser Teil nur illustrativen Charakter hat, liegt damit an folgenden Faktoren:

1. Das Ziel der vorliegenden Untersuchung war es, eine Methodik zu entwickeln, mit deren Hilfe es möglich ist, große Textkorpora nach typischen Mustern im Sprachgebrauch zu untersuchen. Dies wurde in Teil II aufgrund theoretischer Überlegungen in Teil I geleistet. Nun soll im letzten Teil anhand weniger ausgewählter Sprachgebrauchsmuster das Potenzial der Analysemethoden für diskurs- und kulturanalytische Fragestellungen aufgezeigt werden.
2. Zudem liegt der illustrative Charakter der Analysen in Teil III aber auch in der Natur der vorgeschlagenen Methode. Denn sie kann niemals alleine zur Diskursanalyse verwendet werden, sondern nur im Verbund mit traditionellen Zugängen, wie bereits in Kapitel 4.8 deutlich gemacht. Es würde den Rahmen der vorliegenden Untersuchung jedoch sprengen, wenn diese traditionellen Analysemethoden ebenfalls im Detail hergeleitet, begründet und angewandt werden müssten.
3. Es gibt aber noch einen weiteren Grund, weshalb hier keine Diskursanalyse zu einem Thema X geleistet werden soll: Die Chance der vorgeschlagenen Methode einer korpuslinguisti-

schen Diskursanalyse liegt gerade darin, neben Themen, besonders Sprechweisen in den Fokus zu rücken, die unabhängiger von Themen sind. Das Untersuchungsinteresse liegt deshalb weniger in der Suche nach der diskursiven Gestalt von Themen im NZZ-Korpus, sondern in der Suche nach Sprechweisen, die zeit-, textsorten- oder thementypisch sind.

So wird zwar in diesem Teil keine abschließende Diskursanalyse zu einem Thema geleistet, doch es werden einige typische Sprechweisen im NZZ-Korpus genauer analysiert und dadurch gezeigt, wie diese für Beschreibungen von Diskursen genutzt werden können.

## 10 Die Datengrundlage: Das NZZ-Korpus

In Teil II der vorliegenden Untersuchung habe ich bereits mehrfach Untersuchungsergebnisse auf Basis des ‚Neue Zürcher Zeitung-Korpus‘ (NZZ-Korpus 1995–2005) dargestellt. Es umfasst eine Zufallsauswahl von 44 843 Artikeln aus dem Zeitraum von 1995 bis 2005. Weitere wichtige Eckdaten wurden im Rahmen der Überlegungen zur Korpuserstellung (vgl. Kapitel 5.3) bereits genannt. Doch müssen hier weitere Details zu dieser Datengrundlage aufgeführt werden.

### 10.1 Publizistische Daten zur ‚Neuen Zürcher Zeitung‘

Die ‚Neue Zürcher Zeitung‘ ist eine überregionale schweizerische Tageszeitung mit sechs Ausgaben pro Woche (Montag bis Samstag), die 1780 unter dem Titel ‚Zürcher Zeitung‘ gegründet wurde (ab 1821 ‚Neue Zürcher Zeitung‘) und bis heute kontinuierlich erscheint. Die Auflage stieg in der Zeit von 1995 bis 2001 von gut 158 000 auf 170 000 Exemplare und sank danach bis 2005 auf knapp 149 000 verkaufte Exemplare (vgl. Tabelle 10.1 auf der nächsten Seite). Im Vergleich zu anderen kostenpflichtigen Tageszeitungen der Schweiz liegt die Auflage der NZZ damit 2005 nach dem ‚Blick‘ (262 262), dem ‚Tages-Anzeiger‘ (231 182), der ‚Berner Zeitung‘ (227 365) und der ‚Mittelland Zeitung‘ (189 387) an fünfter Stelle.<sup>1</sup> Zusätzlich wird die Presselandschaft seit 1999 durch Gratiszeitungen (als Tageszeitungen), sog. ‚Pendlerzeitungen‘, geprägt. Seit 2004 übertrumpft die Auflage der Gratiszeitung ‚20 Minuten‘ (2005: 380 427 Exemplare) die Auflagen der Kaufzeitungen (vgl. für alle Auflagen von 2005 WEMF AG für Werbemittelforschung 2005).

---

<sup>1</sup> Allerdings handelt es sich bei der ‚Berner Zeitung‘ und der ‚Mittelland Zeitung‘ um sog. ‚Mantelausgaben‘: Ein gemeinsamer internationaler und nationaler Teil (‚Mantel‘) umfasst eine Reihe von relativ eigenständigen Regionalteilen. Diese Regionalredaktionen produzierten ursprünglich eigenständige Lokalzeitungen.

Jahr	verkaufte Auflage
1995	158 167
1996	160 335
1997	162 330
1998	166 525
1999	169 118
2000	169 623
2001	170 113
2002	166 291
2003	159 003
2004	153 025
2005	148 991

**Tabelle 10.1:** Beglaubigte Auflagen der ‚Neuen Zürcher Zeitung‘ (Quellen: AG für die Neue Zürcher Zeitung 2005, 2006 sowie Meldungen in den NZZ-Ausgaben vom 25. 4. 1996, 17. 5. 1997, 22. 4. 1998, 10. 4. 1999, 1. 4. 2000).

Gemäß dem ‚Medienradar‘, einer psychographischen Auswertung von Marktforschungsdaten, lassen sich die Leserinnen und Leser der NZZ schwergewichtig als „progressive Leader“ (statt „konservative Follower“) beschreiben, die zudem eher „materialistisch“ statt „idealistisch“ verortet sind (NZZ Mediadok 2005, 4f., 8).

## 10.2 Eckdaten des Korpus

Für die vorliegende Untersuchung definierte ich alle redaktionellen Artikel der Ausgaben der NZZ von 1995 bis und mit 2005 als Grundgesamtheit (vgl. auch die Ausführungen auf Seite 108). Daraus wurde eine Zufallsstichprobe von 44 843 Artikeln gezogen, die insgesamt einen Umfang von 29,9 Mio. Wörtern umfasst.<sup>2</sup> Diese Stichprobe umfasst damit knapp 6,5 % der 639 213 Artikel der Grundgesamtheit. In Tabelle 10.2 auf der nächsten Seite sind die detaillierten Angaben pro Jahr aufgeführt.

In den folgenden Analysen gebe ich Frequenzen immer als prozentuale Anteile an der Stichprobe (oder Teilen davon, z. B. Jahrgängen) an. Da die Stichprobe repräsentativ für die Grundgesamtheit ist, kann davon ausgegangen werden, dass diese Frequenzangaben auch auf

<sup>2</sup> Die Zufallsauswahl aus allen möglichen Artikeln erledigte ein Perl-Script, das die gewünschte Anzahl Artikelnummern durch einen Zufallsalgorithmus generierte.

Jahr	Grundges. Artikel	Stichprobe Artikel	Wörter	% der GG	Zeitspanne
1995	56 514	3666	2 458 651	6,49	3. Jan. bis 30. Dez. 1995
1996	57 708	3711	2 463 270	6,43	3. Jan. bis 31. Dez. 1996
1997	57 141	3827	2 557 410	6,7	3. Jan. bis 31. Dez. 1997
1998	54 543	3616	2 557 811	6,63	3. Jan. bis 31. Dez. 1998
1999	66 788	4248	2 631 274	6,36	4. Jan. bis 31. Dez. 1999
2000	69 170	4496	2 740 517	6,5	1. Jan. bis 30. Dez. 2000
2001	68 511	4373	2 621 799	6,38	3. Jan. bis 31. Dez. 2001
2002	68 363	4385	2 582 855	6,41	3. Jan. bis 31. Dez. 2002
2003	65 063	4235	2 520 628	6,51	3. Jan. bis 31. Dez. 2003
2004	64 157	4155	2 405 419	6,48	3. Jan. bis 31. Dez. 2004
2005	65 255	4131	2 406 747	6,33	3. Jan. bis 31. Dez. 2005
Total	693 213	44 843	27 946 381	6,47	1995 bis 2005

**Tabelle 10.2:** Datengrundlage der Untersuchung. Artikel der Neuen Zürcher Zeitung zwischen 1995 und 2005 (Zufallsstichprobe, 6 Ausgaben/Woche).

die Grundgesamtheit interpoliert werden können. Die prozentuale Angabe gilt somit auch für die Grundgesamtheit. Neben prozentualen Frequenzangaben nenne ich aber auch hin und wieder die absoluten Frequenzen. Da sich diese absoluten Zahlen auf die Stichprobe beziehen, müssen sie für die Interpolation auf die Grundgesamtheit ungefähr mit dem Faktor 15 multipliziert werden (die Stichprobe umfasst etwa 6,5 % der Artikel der Grundgesamtheit, vgl. Tabelle 10.2).



## 11 Aufbereitung der Daten

Die Artikel wurden mit dem ‚Web as Corpus Toolkit‘ (Ziai/Ott 2005) aus der NZZ-Datenbank heruntergeladen, von HTML-Auszeichnungen gesäubert, tokenisiert und als Text-Dateien abgespeichert.<sup>1</sup> Im Anschluss wurden die Daten in ein Datenbanksystem importiert, in dem zu den Artikeln (unterteilt in ‚Titel‘ und ‚Text‘) folgende Metadaten erfasst sind:

- Zeitungsname
- Publikationsdatum
- Nummer der Ausgabe
- Ressort
- Ressortklasse: Zusammenfassung der Detailressorts zu den Klassen: Ausland, Inland, Wirtschaft, Feuilleton, Lokales, Sport, Vermischtes, Leserbrief, Magazin, übrige, unbekannt
- Seite
- Autor/in
- Anzahl Wörter

Die Textdateien mussten mittels eigens dafür programmierten Filtern ausgelesen werden, um die Metainformationen voneinander zu trennen und in die Datenbank einfügen zu können. Da die originalen Ressortbezeichnungen sehr detailliert sind und teilweise von Jahr zu Jahr ändern, wurden diese zu Ressortklassen zusammengefasst. Da nicht für jeden Artikel Ressortinformationen vorhanden sind (oder sie nicht automatisch extrahiert werden konnten), gibt es eine Gruppe

---

<sup>1</sup> Für die Tokenisierung wurden die Standardeinstellungen des ‚Web as Corpus Toolkits‘ verwendet. Ein Token ist durch ein Leerzeichen oder durch eines der folgenden Zeichen begrenzt: , . ! ? : ; ( ) [ ] { }.



von 2635 Artikeln, für die das Ressort manuell bestimmt werden musste. Tabelle 11.1 auf der nächsten Seite gibt einen Überblick über die Verteilung der Artikel des NZZ-Korpus auf die Ressorts.<sup>2</sup> Daneben zeigt Abbildung 11.1 auf Seite 196 den Umfang der Ressorts pro Jahr in Anzahl Wörtern.

Das Korpus wurde nicht mit morphosyntaktischen Informationen annotiert, da es sich für die vorliegende Untersuchung nicht als

<sup>2</sup> Die Ressortklassen umfassen die folgenden originalen Ressorts (in Klammern: Anzahl Artikel):

**Inland:** Inland (4284), Politische Literatur (126), Eidgenössische Räte (218), Staatspolitisches Forum (18), Eidgenössische Wahlen (14), Kantonsporträt (8), Armee 95 (5), Liechtenstein (1).

**Wirtschaft:** Wirtschaft (6338), (unbekannt) (2819), Börsen und Märkte (1503), Börse Schweiz Text (647), Marktübersicht/Devisen (481), Geld und Anlage (203), Börse Ausland Text (197), Börse New York Text (93), Fokus der Wirtschaft (69), Börsen und Märkte (stocks and markets) (64), Finanzinstrumente (Text) (57), Rohwaren Text (48), Waren-/Finanzmärkte Text (47), Themen und Thesen der Wirtschaft (30), Ökonomische Literatur (28), Anlagefonds (20), Betriebswirtschaft (19), Immobilienmärkte (16), Ausländische Banken in der Schweiz (3), 100 Jahre ASM / SWISSMEM (2), Autozulieferindustrie in der Schweiz (1), Präsenz der Schweizer Wirtschaft im Ausland (1).

**Feuilleton:** Feuilleton (3049), Zürcher Kultur (1205), Literatur und Kunst (346), Film (205), Phono-Spektrum (90), Zeitfragen (64), Kunsthandel/Auktionen (48), Frankfurter Buchmesse (36), Hinweise auf Bücher (35), Kunsthandel-Auktionen (35), Zeitbilder (31), Stimmen zur Zeit und zum Zeitgeschehen (1), Zeit-Bilder seit 1980 (Photo-Reportagen) (1).

**Lokales:** Zürich und Region (3642), Stadt und Kanton Zürich (1758), Zürcher Gemeinderat (82), Zürcher Kantonsrat (81), Wahlen in Zürich (5), Zürcher ZOO (3).

**Sport:** Sport (4612), Schach (109), Olympische Sommerspiele (76), Olympische Spiele (1).

**Vermischtes:** Vermischte Meldungen (2372), Wetter-Vermischte Meldungen (266), Wetter/Vermischtes (137), Wetter/Vermischte Meldg. (134).

**Magazin:** Forschung und Technik (435), Tourismus (349), Medien und Informatik (333), Fernsehen (Text) (218), Philatelie (194), Automobil (115), Lebensart (105), Elektronische Medien (86), Wochenende (84), Mensch und Arbeit (79), Luftfahrt (69), Alpinismus (51), Planen/Bauen/Wohnen (45), Bildung und Erziehung (39), Fernsehen (34), Uhrenmesse (31), Golfsport (30), Schule und Erziehung (30), Technologie und Gesellschaft (21), Mode (15), Bauen/Wohnen (13), Automobilsalon (12), Geschäftsreisen (11), Perspektiven nach dem Studium (11), Orbit (10), Telekommunikation (9), Yachting (8), Radio (6), Blickpunkt (5), Alter (2), Mode Damen (2), Architektur/Design (1).

**Übrige:** Sonderbeilage (35), Kantonsporträt (16), Dokumentation (5), Der Blick auf unsere Leserschaft (3), Expo.02 (3), Sicherheit (2), Das Selbstverständnis der NZZ (1), Das Unternehmen NZZ AG (1), Einkauf Zürich (1), ETH (1), Japan (1).

Die Ressortklasse ‚Ausland‘ lautet gleich wie das originale Ressort. ‚Leserbriefe‘ lautet im Original ‚Briefe an die NZZ‘. Die 2635 manuell einer Ressortklasse zugeordneten Artikel wurden nicht einem der Originalressorts zugeteilt.

Ressort	Anzahl Artikel	Anzahl Wörter
Ausland	5670	3 413 875
Inland	4878	3 050 673
Wirtschaft	10 654	5 538 724
Feuilleton	5543	4 699 157
Lokales	5743	2 710 816
Sport	5153	4 437 624
Vermischtes	3035	1 024 520
Magazin	2947	2 470 711
Leserbriefe	1022	344 507
übrige	122	186 053
(unbekannt)	76	69 721
Total	44 843	27 946 381

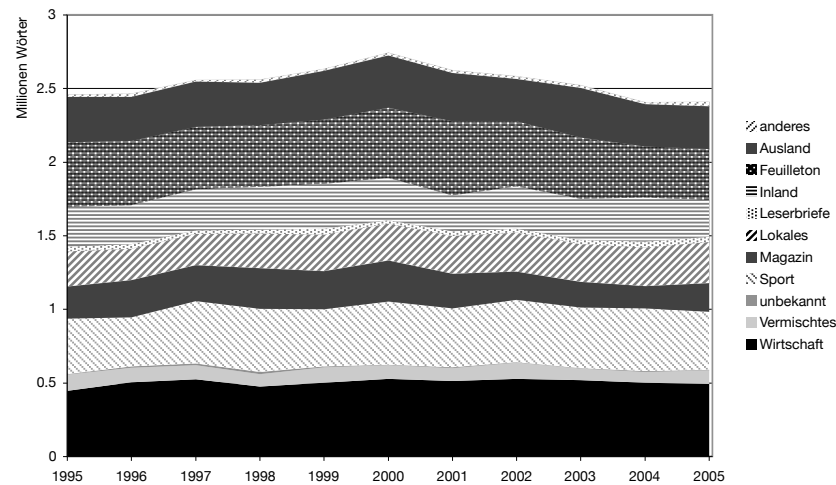
**Tabelle 11.1:** Ressorts im NZZ-Korpus.

zwingend notwendig erwies.<sup>3</sup> Hingegen zeigte sich im Verlauf der Analysen, dass Auszeichnungen mit geografischen Informationen im Korpus hilfreich wären. Diese habe ich zwar nicht als eigentliche Annotation vorgenommen, sondern verwendete Listen von Staaten, Hauptstädten und Kontinenten der Welt, um automatisiert danach suchen zu können. Die Listen enthalten jeweils nicht nur Nomina, sondern auch die adjektivischen Bezeichnungen, sowie die Bezeichnungen für die Bewohnerinnen und Bewohner der Länder.<sup>4</sup> Unter Verwendung des ‚Geocoding‘-Services von ‚Google‘<sup>5</sup> wurden die Orte mit Längen- und Breitengraden referenziert, sodass sie auf einer Karte automatisiert aufgetragen werden können.

<sup>3</sup> Vgl. dazu auch die Überlegungen zur Lemmatisierung und dem Tagging in Kapitel 6.5. Zudem hätten sich einige technische Probleme ergeben: Der Tagger hätte für schweizerisches Standarddeutsch trainiert (z. B. ‚ss‘ statt ‚ß‘-Schreibung) und die inkonsistente Behandlung von Sonderzeichen (teilweise ausgeschriebene Umlaute) beseitigt werden müssen. Das sind zwar technisch lösbare Probleme, angesichts des angezweiferten Nutzens des Taggings für die Analyse verzichtete ich aber darauf. Ich werde aber im Schlusskapitel (vgl. Kapitel 17.1) der vorliegenden Arbeit im Sinne eines Ausblicks Analysemöglichkeiten in einem Korpus skizzieren, das lemmatisiert und getagged ist.

<sup>4</sup> Die manuell erstellten Listen basieren auf dem Verzeichnis „Verzeichnis der Staatennamen für den amtlichen Gebrauch in der Bundesrepublik Deutschland“ (Sprachendienst des Auswärtigen Amts 2006) sowie der „Liste der Hauptstädte der Welt“ der „freien Enzyklopädie Wikipedia“ (Freie Enzyklopädie Wikipedia o. J.). Sie wurden erweitert um inoffizielle, aber gebräuchliche Bezeichnungen wie *Libyen* oder *USA* oder um nicht mehr existierende Bezeichnungen wie *Jugoslawien*.

<sup>5</sup> Vgl. <http://code.google.com/apis/maps/> (14. 2. 2008).



**Abbildung 11.1:** Umfang der Ressorts pro Jahr im NZZ-Korpus in Anzahl Wörter.

Die elektronischen Daten im NZZ-Archiv sind bezüglich der Schreibung von Sonderzeichen inkonsistent. In den früheren Jahren wurden aus technischen Gründen alle Umlaute in ausgeschriebene Varianten konvertiert („ue“ statt „ü“ etc.). Diese Inkonsistenz ist nun auch im NZZ-Korpus vorhanden und musste bei Suchanfragen und Berechnungen berücksichtigt werden.<sup>6</sup> Sofern in den Analysen nichts Gegenteiliges vermerkt ist, wurden jeweils immer beide Schreibvarianten eines Suchausdrucks mit Umlaut gesucht. In den aufgeführten Belegen habe ich aber alle Umlautumschreibungen wieder auf die ursprünglich in der Zeitung abgedruckten Varianten zurückgeführt.

In Kapitel 9.2 wird das Datenbanksystem, das ich zur Verwaltung und Abfrage des NZZ-Korpus verwende, detailliert vorgestellt.

<sup>6</sup> So werden auch bei der Berechnung der für ein Teilkorpus signifikanten Mehrworteinheiten die Schreibvarianten jeweils auf eine zurückgeführt, um die Resultate durch die Varianten nicht zu verfälschen.

## 12 Berechnung der typischen Mehrworteinheiten

### 12.1 Festlegung von Parametern und Teilkorpora

Um in den Korpusdaten die frequenten und/oder statistisch signifikanten Mehrworteinheiten zu berechnen, verwendete ich die Software ‚Ngram Statistics Package‘ (‚NSP‘) (Banerjee/Pedersen 2003), das sich nach der Evaluation verschiedener Programme als für meine Zwecke beste Lösung erwies (vgl. Kapitel 9.3). Allerdings ergaben sich angesichts der umfangreichen Datenmengen Leistungsprobleme mit NSP. Das machte es nötig, die Software zu modifizieren, um solche Datenmengen verarbeiten zu können.<sup>1</sup>

Mit NSP wurden für verschiedene Teilkorpora die frequenten Mehrworteinheiten berechnet. Die Teilkorpora orientieren sich entlang den Kategorien ‚Zeit‘ und ‚Ressort‘. Es wurden drei verschiedene zeitliche Raster verwendet, um die Balance zwischen Auflösungsgrad und Mindestartikelmenge pro Teilkorpus zu erreichen. Das erste Raster fasst einzelne Jahre zusammen, das zweite fasst den ganzen Zeitraum von 1995–2005 als Einheit auf und im dritten Raster werden zwei Einheiten à drei Jahre gebildet. Für die zweite Dimension, die Ressorts, wurden acht der elf definierten Ressortklassen gewählt. Die Klassen ‚anderes‘, ‚Magazin‘ und ‚unbekannt‘ wurden ignoriert, da sie sehr heterogen zusammengesetzt sind und darüber hinaus relativ wenige Artikel enthalten. So sind folgende Teilkorpora definiert:

1. Für jedes Jahr von 1995–2005 je ein Teilkorpus pro Ressort
  - a) Ausland
  - b) Inland

---

<sup>1</sup> Im Detail: Ein Perlscript sorgt dafür, dass die Daten in kleinere Einheiten aufgeteilt und je separat von NSP verarbeitet werden. Zudem nutzt das Script Mehrprozessorsysteme besser aus, indem die Teilaufgaben auf mehrere ‚Threads‘ aufgeteilt und so parallel verarbeitet werden können. Nach Abschluss der Berechnungen werden die Resultate der kleineren Einheiten wieder zusammengefügt. Das Script kann beim Autor bezogen werden: [www.bubenhofer.com/korpusanalyse/](http://www.bubenhofer.com/korpusanalyse/).

- c) Wirtschaft
  - d) Feuilleton
  - e) Vermischtes
  - f) Sport
  - g) Lokales
  - h) Leserbriefe
2. Zeitraum 1998–2005 pro Ressort
    - a) Ausland
    - b) ...
  3. Zeitraum 1995–1997 pro Ressort
    - a) Ausland
    - b) ...
  4. Zeitraum 2003–2005 pro Ressort
    - a) Ausland
    - b) ...

Für die Berechnung der Mehrworteinheiten müssen eine Reihe von Parametern gesetzt werden wie Länge der Mehrworteinheit oder maximale Spannweite zwischen erstem und letztem Wort der Mehrworteinheit (vgl. Ausführungen auf Seite 117). Als Standardverfahren haben sich folgende Parameter als gewinnbringend für die vorliegenden Analysezwecke erwiesen:

*Länge:* 3 Wörter

*Spannweite:* 10 Wörter (max. Distanz zwischen 1. und letztem Wort)

*Basis:* Token (nicht Lemmata)<sup>2</sup>

*Mindestfrequenz:* 6 pro Mio. Wörter

---

<sup>2</sup> Vgl. dazu auch Kapitel 6.5.

Die Mindestfrequenz wurde relativ zur Korpusgröße definiert, kann aber aufgrund der Berechnungsmethode leicht variieren.<sup>3</sup> In den Resultaten weiter unten gelten jeweils diese Parameter des Standardverfahrens. Die Parameter werden nur bei davon abweichenden Berechnungen speziell erwähnt.

Im Anschluss wurden die Listen der für ein Teilkorpus berechneten Mehrworteinheiten durch Kontrastverfahren untereinander verglichen, um die für die Teilkorpora typischen Mehrworteinheiten zu finden (vgl. Kapitel 8.1). Dafür diente ein eigenes Programm, das jeweils zwei Listen von Mehrworteinheiten wahlweise mittels der Statistiken ‚Log-Likelihood Koeffizient‘, ‚ $\chi^2$ -Test‘ oder ‚Mann-Whitney-Rank-Test‘ vergleicht.<sup>4</sup> Das Programm ermöglicht aber nicht nur den Vergleich zweier Listen miteinander, sondern den Vergleich einer mit mehreren Listen. So können die für ein Teilkorpus im Vergleich zu allen anderen Teilkorpora typischen Mehrwortgruppen eruiert werden.

So wurden für jedes Teilkorpus die im Vergleich mit den anderen Teilkorpora typischen Mehrworteinheiten für ein Typikprofil berechnet. Aus Platzgründen wird hier nur ein Teil dieser Listen – und von diesen jeweils nur die nach dem Log-Likelihood-Koeffizienten signifikantesten Mehrworteinheiten – wiedergegeben.<sup>5</sup> Darüber hinaus sind die hier aufgeführten Listen manuell gefiltert worden. Dabei wurden Mehrworteinheiten entfernt, die folgende Elemente enthalten:

1. Metainformationen, die nicht zum eigentlichen Artikel gehören, wie Datenbankfeldbezeichnungen, Paginierung, Zeitungstitel etc.
2. ausschließlich Zahlen

<sup>3</sup> Da die Teilkorpora aus Performance-Gründen für die Berechnung in kleinere Teile separiert werden müssen (vgl. Fußnote 1 auf Seite 197), ergeben sich bei den Mindestfrequenzen unter Umständen Abweichungen: Die Teile sind höchstens 50 000 Wörter groß, können aber bei ungünstigen Teilungsverhältnissen auch kleiner sein. Trotzdem ist die geforderte Mindestfrequenz der Mehrworteinheiten pro Teil 3 (hochgerechnet auf 1 Mio. Wörter also 6). Wegen der Aufteilung kann aber nicht ausgeschlossen werden, dass bei einer anderen Teilung gewisse Mehrworteinheiten die Mindestfrequenzmarke über- oder unterschritten hätten und damit leicht andere Mehrwortlisten berechnet würden. Diese Ungenauigkeiten können angesichts der Datenmengen m. E. aber ignoriert werden.

<sup>4</sup> Vgl. für die Details dieser statistischen Verfahren Kapitel 7.2.3 bis 7.2.5.

<sup>5</sup> Alle Listen in vollständiger Form sind online abrufbar: <http://www.bubenhofner.com/korpusanalyse/>.

### 3. Korrespondenten-/Autorinnen-Kürzel und -Namen

Tabelle 12.1 auf der nächsten Seite zeigt die ungefilterte Liste an Mehrworteinheiten, die beim Vergleich des Teilkorpus ‚Ausland 1995–1997‘ mit dem Teilkorpus ‚Ausland 2003–2005‘ zu Stande gekommen ist. Die Mehrworteinheiten sind nach dem Log-Likelihood-Koeffizienten  $G^2$  absteigend sortiert; an der Spitze befinden sich also die für den ersten Zeitraum statistisch typischsten Mehrworteinheiten. Die weiteren Spalten zeigen die absoluten und relativen (pro Million Wörter gerechneten) Frequenzen der Mehrworteinheit in den beiden Korpora.

Der erste Blick auf die Tabellen ist ernüchternd. Auf den ersten neun Zeilen sind verschiedene Kombinationen von Metadaten sichtbar, die aus Datums- und Seitenangaben bestehen. Es folgen verschiedene Varianten der Agenturbezeichnung *afp Agence France Presse*. Daran schließen die für die Analyse interessanten Mehrworteinheiten an.

Am Ende dieser Liste, von der in Tabelle 12.1 auf der nächsten Seite nur die ersten 43 Zeilen abgedruckt sind, finden sich mit negativen  $G^2$ -Werten die für das Vergleichskorpus ‚Ausland 2003–2005‘ typischen Mehrworteinheiten (als Tabelle 12.2 auf Seite 202 ab der Zeile 60704 abgedruckt). Die ganze Liste des Vergleichs umfasst also 60746 Zeilen.

#### 12.2 Selektion der Mehrworteinheiten für die Analyse

Die in Tabellen 12.1 auf der nächsten Seite und 12.2 auf Seite 202 dargestellten Mehrworteinheiten, die typisch für die Zeiträume 1995–1997 bzw. 2003–2005 sind, umfassen gerade einmal 1,4‰ der Mehrworteinheiten des kompletten Typikprofils. Nur einige dieser Mehrworteinheiten scheinen für eine weitere Analyse interessant und es ist klar, dass diese Listen stark gefiltert werden können. Die Listen enthalten viele Metadaten, die für die weitere Analyse ignoriert werden können. Weiter können bestimmte Mehrworteinheiten zusammengefasst werden. So scheint es sinnvoll, Mehrworteinheiten, die aus zwei gemeinsamen Wörtern bestehen, als Cluster zu markieren.<sup>6</sup> In Tabelle 12.2 auf Seite 202 würden *der im Irak* unter *die im Irak* mit dem stärkeren Log-Likelihood-Wert subsumiert werden, nicht jedoch *in*

<sup>6</sup> Bei einem Cluster werden in den Tabellen die jeweils der ersten Zeile folgenden Zeilen leicht eingerückt, um die Zusammengehörigkeit visuell zu betonen.

Zeile	G <sup>2</sup>	MWE	# absolut		#/Mio.	
			95-97	02-05	95-97	02-05
1	591,42	1996 date S	449	0	45,38	0,00
2	586,15	1997 date S	445	0	44,97	0,00
3	496,58	date S 2	377	0	38,10	0,00
4	395,16	1995 date S	300	0	30,32	0,00
5	330,61	date S 3	251	0	25,37	0,00
6	275,29	date S 1	209	0	21,12	0,00
7	264,75	date S 5	201	0	20,31	0,00
8	89,57	date S 9	68	0	6,87	0,00
9	85,62	date S 7	65	0	6,57	0,00
10	55,32	Agence France Presse	42	0	4,24	0,00
11	55,32	Ausland Agence Presse	42	0	4,24	0,00
12	55,32	afp France Presse	42	0	4,24	0,00
13	55,32	afp Agence Presse	42	0	4,24	0,00
14	55,32	Ausland France Presse	42	0	4,24	0,00
15	55,32	Ausland afp Presse	42	0	4,24	0,00
16	48,69	die in des	61	6	6,17	0,65
17	47,15	der bosnischen Serben	68	9	6,87	0,98
18	42,15	Human Rights Watch	32	0	3,23	0,00
19	40,83	in der bosnischen	31	0	3,13	0,00
20	39,52	und der Opposition	30	0	3,03	0,00
21	36,88	der ist und	28	0	2,83	0,00
22	36,19	der der Partei	90	24	9,10	2,60
23	35,56	Aung San Suu	27	0	2,73	0,00
24	34,25	in der ersten	26	0	2,63	0,00
25	32,93	der der Menschenrechte	25	0	2,53	0,00
26	31,71	der in Bosnien	40	4	4,04	0,43
27	31,61	der der sei	24	0	2,43	0,00
28	31,61	sowie die und	24	0	2,43	0,00
29	31,61	der der ersten	24	0	2,43	0,00
30	31,61	durch die des	24	0	2,43	0,00
31	31,61	haben am Donnerstag	24	0	2,43	0,00
32	30,30	am abend in	23	0	2,32	0,00
33	30,30	San Suu Kyi	23	0	2,32	0,00
34	30,30	von der dem	23	0	2,32	0,00
35	30,30	Aung San Kyi	23	0	2,32	0,00
36	30,30	Aung Suu Kyi	23	0	2,32	0,00
37	30,30	der ersten Runde	23	0	2,32	0,00
38	29,48	der der bosnischen	35	3	3,54	0,33
39	28,98	ist nicht zu	22	0	2,22	0,00
40	28,98	in der Serben	22	0	2,22	0,00
41	28,98	mit der an	22	0	2,22	0,00
42	28,98	auch an der	22	0	2,22	0,00
43	27,90	die bosnischen Serben	42	6	4,24	0,65

**Tabelle 12.1:** Typische Mehrworteinheiten im Ressort ‚Ausland‘ des NZZ-Korpus in den Jahren 1995–1997 im Vergleich zu den Jahren 2003–2005.



Zeile	G <sup>2</sup>	MWE	# absolut		#/Mio.	
			95-97	02-05	95-97	02-05
60704	-68,52	afp Agence Press	0	47	0,00	5,10
60705	-68,52	Ausland Agence Press	0	47	0,00	5,10
60706	-68,52	afp France Press	0	47	0,00	5,10
60707	-68,52	Ausland afp Press	0	47	0,00	5,10
60708	-68,52	Agence France Press	0	47	0,00	5,10
60709	-68,52	Ausland France Press	0	47	0,00	5,10
60710	-68,52	in den Irak	0	47	0,00	5,10
60711	-71,51	p Neue Zürcher	1209	1556	122,19	168,72
60712	-72,52	p Neue Zeitung	1206	1556	121,89	168,72
60713	-72,86	Neue Zeitung date	1205	1556	121,79	168,72
60714	-72,86	Zürcher Zeitung date	1205	1556	121,79	168,72
60715	-72,86	Neue Zürcher date	1205	1556	121,79	168,72
60716	-72,86	p Zürcher Zeitung	1205	1556	121,79	168,72
60717	-72,89	Ausland AR A	0	50	0,00	5,42
60718	-72,89	Ausland AR Rüesch	0	50	0,00	5,42
60719	-72,89	S AR Rüesch	0	50	0,00	5,42
60720	-72,89	S Ausland Rüesch	0	50	0,00	5,42
60721	-72,89	S Ausland AR	0	50	0,00	5,42
60722	-72,89	im Irak die	0	50	0,00	5,42
60723	-72,89	AR Rüesch A	0	50	0,00	5,42
60724	-72,89	Ausland Rüesch A	0	50	0,00	5,42
60725	-86,02	S 2 Schweiz	0	59	0,00	6,40
60726	-86,02	S 2 sda	0	59	0,00	6,40
60727	-86,02	2 Ausland Schweiz	0	59	0,00	6,40
60728	-86,02	2 sda Schweiz	0	59	0,00	6,40
60729	-86,02	2 Ausland sda	0	59	0,00	6,40
60730	-91,79	2 Ausland Press	47	179	4,75	19,41
60731	-99,94	Ausland Associated Press	104	284	10,51	30,80
60732	-99,94	ap Associated Press	104	284	10,51	30,80
60733	-99,94	S ap Associated	104	284	10,51	30,80
60734	-99,94	S Ausland Associated	104	284	10,51	30,80
60735	-99,94	Ausland ap Associated	104	284	10,51	30,80
60736	-99,94	Ausland ap Press	104	284	10,51	30,80
60737	-99,94	S Ausland ap	104	284	10,51	30,80
60738	-153,08	die im Irak	0	105	0,00	11,39
60739	-164,74	der im Irak	0	113	0,00	12,25
60740	-167,66	S Ausland Schweiz	0	115	0,00	12,47
60741	-167,66	S Ausland sda	0	115	0,00	12,47
60742	-167,66	S sda Schweiz	0	115	0,00	12,47
60743	-167,66	Ausland sda Schweiz	0	115	0,00	12,47
60744	-705,63	2005 date Nr	0	484	0,00	52,48
60745	-740,62	2004 date Nr	0	508	0,00	55,08
60746	-814,98	2003 date Nr	0	559	0,00	60,61

**Tabelle 12.2:** Typische Mehrworteinheiten im Ressort ‚Ausland‘ des NZZ-Korpus in den Jahren 2003–2005 im Vergleich zu den Jahren 1995–1997.

*den Irak* und *im Irak die*. Wenn nach der Regel der Übereinstimmung mit zwei von drei Wörtern Cluster gebildet werden, finden sich unter diesen Clustern Mehrworteinheiten, die auch in ihrer Kombination im Korpus auftauchen können. Ein Beispiel dafür ist *San Suu Kyi* und *Aung San Suu*, die auf *Aung San Suu Kyi* zurück gehen (vgl. Tabelle A.1 auf Seite 341, Zeilen 14 und 15). Dieser Zusammenhang ist jedoch keineswegs zwingend, wie das Beispiel mit *die im Irak* und *der im Irak* zeigt. Dafür wird bei letzterem Beispiel durch das Clusterverfahren eine strukturelle Ähnlichkeit sichtbar.

Doch auch wenn die Mehrworteinheiten, die Metadaten anzeigen, ignoriert und ähnliche Mehrworteinheiten als Cluster zusammengeführt werden, sind die Typikprofile noch immer sehr umfangreich. Je nach Forschungsinteresse wird man sich auf bestimmte Mehrworteinheiten beschränken. Dafür ist eine Klassifikation der Mehrworteinheiten notwendig. Diese Klassifikation ist ihrerseits von den Forschungsinteressen abhängig, trotzdem kann folgende Systematik, nach der ich in der vorliegenden Arbeit vorgegangen bin, auch für anders gelagerte Analysen nützlich sein.

- Selektionskriterium 1: Statistische Typik
  1. Mehrworteinheit ist typisch für bestimmte Teilkorpora
  2. Mehrworteinheit ist über alle Teilkorpora ähnlich verteilt
- Selektionskriterium 2: Syntaktische Festigkeit
  1. Tendenziell syntaktisch fest
  2. Tendenziell syntaktisch variabel
- Typologisierendes Kriterium 1: Kontinuität der Wortfolge
  1. Tendenziell diskontinuierliche Mehrworteinheiten
  2. Tendenziell kontinuierliche Mehrworteinheiten
- Typologisierendes Kriterium 2: Phraseologischer Typus
  1. Referentielle Phraseologismen
  2. Strukturelle Phraseologismen
  3. Kommunikative Phraseologismen
  4. Sonderformen

Das primäre Selektionskriterium ist die statistische Typik der Mehrworteinheit für ein bestimmtes Teilkorpus. Einerseits können Einheiten in den Blick genommen werden, die besonders typisch für einen Zeitabschnitt, eine Textsorte, ein Ressort etc. sind, oder aber solche, die es gerade nicht sind. Ist man an grundsätzlichen sprachlichen Mustern interessiert, die unabhängig von Zeit und Textsorte vorkommen, würde man die Mehrworteinheiten auswählen, die über das ganze Korpus oder mehrere Korpora ähnlich verteilt sind.

In der vorliegenden Untersuchung interessiere ich mich aber primär für Mehrworteinheiten, die typisch für einen Zeitabschnitt oder ein Ressort sind. Deshalb wurden ja auch Typikprofile erstellt, die pro Ressort die zeittypischen Mehrworteinheiten umfassen. Für die Untersuchung wähle ich also aus den Typikprofilen jene Mehrworteinheiten aus, die besonders signifikant für das Teilkorpus sind. Da die Typikprofile, die ich im folgenden in den Beispielanalysen verwende, jeweils nur die Zeitabschnitte 1995–1997 und 2003–2005 entgegenstellen, muss jeweils geprüft werden, ob sich auch über die ganze Periode des Untersuchungskorpus von 1995–2005 eine Verteilung mit signifikanten Unterschieden ergibt. Wenn das nicht der Fall ist, wie z. B. beim Ausdruck *Kampf gegen*, gilt das vielleicht für einen verwandten Ausdruck, der zum selben Cluster gehört, wie z. B. *Kampf gegen ... Terror*.

Als sekundäres Selektionskriterium gilt die syntaktische Festigkeit: Die Mehrworteinheiten im Typikprofil werden zunächst danach untersucht, ob sie syntaktisch eher fest oder eher variabel sind. Eine Mehrworteinheit ist dann syntaktisch fest, wenn der größte Teil der Belege für die Mehrworteinheit dem selben syntaktischen Muster folgen. Bei *die die Nato*<sup>7</sup> wird das kaum der Fall sein: Die beiden Artikel *die* können je in unterschiedlichen Gliedsätzen stehen, die in jeweils unterschiedlichen syntaktischen Bezügen zueinander stehen. Anders bei der Mehrworteinheit *gegen der Regierung*, das in den meisten Fällen nach folgendem syntaktischen Muster *gegen X<sub>AKK.-OBJ.</sub> der Regierung<sub>GEN.-OBJ.</sub>* verwendet wird:

- (17) Etwa 18 000 Arbeiter sind in Bukarest auf die Strasse gegangen, um **gegen** die Wirtschaftspolitik **der Regierung** zu pro-

<sup>7</sup> Alle folgenden Beispiele stammen aus dem Typikprofil A.1.1 auf Seite 341.

testieren. *Neue Zürcher Zeitung* vom 13. April 1995, Ressort ‚Ausland‘, ap: „Arbeiterdemonstration in Bukarest“

- (18) Anguita, der sich wie schon zu Zeiten gemeinsamer Opposition auf persönlicher Ebene bestens mit dem konservativen Regierungschef zu verstehen scheint, kündigte entschlossenen Widerstand **gegen** die „rechte sowie neoliberale Politik“ **der Regierung** an, einschliesslich Massenmobilisierungen und Generalstreiks. *Neue Zürcher Zeitung* vom 27. Juli 1996, Ressort ‚Ausland‘, Serna, A.: „Spaniens Spitzenpolitiker proben den Dialog. Mit Spannung erwartetes Treffen von Gonzalez und Anguita“

Syntaktisch feste Mehrworteinheiten sind für die weiteren Analysen interessanter als *variable*, da sie auf Verwendungskontexte zurückgehen, die zumindest syntaktisch aber wahrscheinlich auch semantisch ähnlich sind. Syntaktisch *variable* Dreiworteinheiten umfassen neben einem eher niedrigfrequenten Lexem meist zwei hochfrequente Wörter wie Artikel oder Präpositionen. In der Kombination von niedrig- und hochfrequentem Wortmaterial ergeben sich zwar statistisch signifikante Wortverbindungen, die aber eben auf syntaktisch völlig unterschiedliche Muster zurückgehen.

Das Beispiel *gegen ... der Regierung* oben hat bereits gezeigt, dass es Mehrworteinheiten gibt, die syntaktisch fest sind, aber deren einzelne Wörter nicht unmittelbar aufeinander folgen (‚diskontinuierliche Mehrworteinheiten‘). Zwischen den Wörtern der Einheit sind Lücken, ‚Slots‘, die durch Einzelwörter, unvollständige oder komplette Satzglieder gefüllt werden. Diese ‚Füllungen‘ der Slots sind variabel im Bezug auf die sie aufnehmende Mehrworteinheit. Gerade bei diskursanalytischen Fragestellungen sind diese Füllungen besonders interessant zu untersuchen, da es anscheinend Faktoren gibt, die die Art der Füllung der Slots steuern.

Für die meisten Forschungsfragen wird die syntaktische Festigkeit das wichtigste Kriterium sein, um eine Auswahl von Mehrworteinheiten für die weitere Analyse zu treffen. Für das Kriterium der Slots (‚kontinuierlich‘ vs. ‚diskontinuierlich‘) wird das bereits nicht mehr gelten, so habe ich in der vorliegenden Untersuchung sowohl Mehr-

worteinheiten mit als auch solche ohne Slots für die Detailanalyse hinzugezogen.

Werden syntaktisch variable Mehrworteinheiten ignoriert, können die verbleibenden Einheiten nach weiteren Kriterien klassifiziert werden. Nützlich ist beispielsweise eine Klassifikation, wie sie Burger (1998, 35ff.) für Phraseologismen vorschlägt. Um das ganze Spektrum von nur sehr schwach oder gar nicht idiomatischen bis zu hochidiomatischen Wortverbindungen zu klassifizieren, unterscheidet er grob zwischen referentiellen, strukturellen und kommunikativen Phraseologismen. ‚Referentielle Phraseologismen‘ „beziehen sich auf Objekte, Vorgänge oder Sachverhalte der Wirklichkeit“: *Schwarzes Brett, jmdm. übers Ohr hauen, Morgenstern hat Gold im Mund*. ‚Strukturelle Phraseologismen‘ „haben ‚nur‘ eine Funktion innerhalb der Sprache, nämlich die Funktion, (grammatische) Relationen herzustellen“: *in Bezug auf, sowohl – als auch*. ‚Kommunikative Phraseologismen‘ oder ‚Routineformeln‘ „haben bestimmte Aufgaben bei der Herstellung, Definition, dem Vollzug und der Beendigung kommunikativer Handlungen“: *Guten Morgen, ich meine* (Burger 1998, 36). Aus phraseologischer Sicht ist die Gruppe der strukturellen Phraseologismen die „am wenigsten interessante“ (Burger 1998, 36), aus diskurs- oder kulturalistischer hingegen lohnt ein detaillierter Blick auf solche Formeln sehr wohl. Neben diesen Basisklassen werden aber auch einige Spezialformen verwendet, die sich in den Typikprofilen der vorliegenden Arbeit ebenfalls finden. So z. B. ‚onymische Phraseologismen‘, die die Funktion von Eigennamen haben (*das Rote Kreuz, der Ferne Osten, das Weiße Haus*) oder ‚phraseologische Termini‘ (*rechtliches Gehör, einstweilige Verfügung, in Konkurs gehen*), die innerhalb eines fachlichen Subsystems normiert sind (Burger 1998, 46f.).

Arbeitet man mit zu Clustern zusammengefassten Mehrworteinheiten und ignoriert jene, die syntaktisch eher variabel sind, wird ein Typikprofil, wie ich es in der vorliegenden Arbeit für die Zeitungsressorts erstellt habe, gut überblickbar. Konzentriert man sich in der Folge auf entweder die für eine Zeitperiode besonders typischen Mehrworteinheiten, reduziert sich die Menge weiter. Am Beispiel des Typikprofils des Auslandsressorts ergeben sich folgende Mengen: Das gesamte Typikprofil aller Dreiworteinheiten von 1995–1997 und 2003–2005 (berechnet nach den Parametern, wie sie auf Seite 198 dar-

gelegt sind) umfasst 60 745 Mehrworteinheiten. Konzentriert man sich auf die jeweils für einen der Zeiträume signifikantesten Mehrworteinheiten ( $G^2 \geq 10$ ), ergibt das 994 bzw. 1152 Mehrworteinheiten für den älteren/jüngeren Zeitraum. Jeweils etwa 10 % dieser Mengen können ignoriert werden, da sie Metainformationen enthalten. Durch das Clustering der Mehrworteinheiten kann die Übersicht verbessert werden, da syntaktisch ähnliche Einheiten zusammengefasst werden. Für diskursanalytische Fragestellungen, die den Sprachgebrauch im Blick haben, müssen alle diese Mehrworteinheiten analysiert werden, wobei bei vielen eine rasche Korpusrecherche zeigt, ob sich eine weitere Analyse lohnt. So wird oft ersichtlich, dass die Unterschiede der Distribution über das Korpus doch nicht signifikant sind oder dass der Ausdruck auf eine mehrfach wörtlich zitierte größere Textpassage zurück geht und deshalb für die Fragestellung nicht relevant ist.

Bei thematisch orientierten Diskursanalysen oder bei kulturanalytischen Fragestellungen, die beispielsweise ganz bestimmte Formen des Sprachgebrauchs (z. B. Personennennungen, Struktur von semantischen Feldern, Valenz von Verben etc.) untersuchen wollen, würde das Typikprofil weiter gefiltert werden. So fallen unter den für das Ressort ‚Ausland‘ typischen Mehrworteinheiten solche auf, die Nomen oder Verben aus den Wortfeldern KRIEG, KAMPF und TERROR enthalten. Für diskursanalytische Fragestellungen in diesem Bereich ist es naheliegend, Mehrworteinheiten mit einer entsprechenden Semantik für die Untersuchung auszuwählen. Eine andere Systematik der Selektion ergibt sich aber auch, wenn man weniger an Sprachgebrauch interessiert ist, der besonders typisch für einen Bereich ist, sondern allgemeinere Muster im Blick hat. Dann würden nicht die für die Teilkorpora signifikantesten, sondern im Gegenteil die nicht-signifikanten Mehrworteinheiten ausgewählt.

Für die vorliegende Arbeit habe ich Mehrworteinheiten ausgewählt, die

1. syntaktisch eher fest sind,
2. als strukturelle Phraseologismen variable, für Teilkorpora typische Slots aufweisen oder
3. als referentielle oder kommunikative Phraseologismen besonders typisch für bestimmte Teilkorpora sind.

Die Krux an diesen Selektionskriterien ist, dass oft eine minimale Korpusrecherche nötig ist um abzuklären, ob die Mehrworteinheit unter eines der Kriterien fällt. Es ist klar, dass z. B. eine syntaktische Annotation des Korpus oder die Berechnung der Varianz der Distanzen zwischen den Token der Worteinheit diesbezüglich hilfreich wären.

Im Anhang in Kapitel A finden sich Tabellen mit den typischen Mehrworteinheiten, die manuell bereinigt und zu Clustern zusammengefasst wurden. Sie sind nach verschiedenen Ressorts und den Zeitabschnitten 1995–1997 und 2003–2005 gegliedert. Auch diese Tabellen stellen jeweils nur die signifikantesten Mehrworteinheiten dar. In elektronischer Form können aber die vollständigen Listen aller Ressorts und aller Zeitabschnitte der drei Raster (Jahres-, Dreijahres- und Elfjahresraster) abgerufen werden.<sup>8</sup>

In den folgenden Kapiteln zeige ich jeweils gegenüber den Listen im Anhang (Kapitel A) weiter manuell gefilterte Listen von Mehrworteinheiten, aus denen eine Auswahl für exemplarische Analysen hinzugezogen werden.

---

<sup>8</sup> Vgl. <http://www.bubenhofer.com/korpusanalyse/>.

## 13 Beispielanalysen

### 13.1 Die Auslandsberichterstattung im NZZ-Korpus

Das Ressort ‚Ausland‘ enthält in der NZZ nach ‚Wirtschaft‘ (inkl. ‚Börse‘) am meisten Artikel und Wörter.<sup>1</sup> Als Ausgangspunkt für die Analyse in der vorliegenden Arbeit nutze ich die berechneten Mehrwortlisten, die die Mehrworteinheiten auflisten, welche typisch für die Zeiträume 1995–1997 und 2003–2005 sind. Die Tabellen A.1 auf Seite 341 und A.2 auf Seite 343 im Anhang zeigen manuell gefilterte und geclusterte Listen mit Mehrworteinheiten, die für die jeweilige Periode signifikanter als  $G^2 \geq 15$  sind. Daraus entnehme ich 36 bzw. 35 Mehrworteinheiten, die mir für die weitere Analyse besonders interessant erscheinen (vgl. Tabelle 13.1 auf der nächsten Seite).

#### 13.1.1 Grobanalyse

Im ersten Überblick, und unter Einbezug von Weltwissen, wird ersichtlich, dass ein Teil der Mehrworteinheiten in Tabelle 13.1 auf der nächsten Seite offenbar in Abhängigkeit zu bestimmten Ereignissen dieser Zeitperioden steht. So ist für die Periode 2003–2005 der *Kampf gegen Terrorismus* (Zeile 44ff.)<sup>2</sup> Hintergrund des *Kriegs im Irak* (36f.) mit den Akteuren *Präsident Bush* (52), *Usama bin Ladin* (59) und *Saddam Hussein* (66). Wobei offensichtlich ein möglicher *Abzug aus dem Irak* (42) thematisiert wird – vielleicht wegen der hohen *Zahl der* (67) *Opfer* (65). Diese Ausdrücke weisen jeweils eine Vielzahl weiterer Kollokatoren auf. Etwa 100 unterschiedliche Mehrworteinheiten, die Teile des Namens von *Saddam Hussein* enthalten, ergeben Kollokatoren wie: *von/auf/gegen (Saddam Hussein)*, *Diktator*, *Kopfgeld* ... *ausgesetzt*, *Regime*, *(Saddam) mit Hilfe*, *Saddam-Statue beschlag-*

---

<sup>1</sup> Vgl. dazu Tabelle 11.1 auf Seite 195.

<sup>2</sup> Zahlen in Klammern beziehen sich im Folgenden auf die Zeilennummer der vorher genannten Tabelle.



	G <sup>2</sup>	MWE 1995–1997		G <sup>2</sup>	MWE 2003–2005
1	47,15	der bosnischen Serben	36	–72,89	im Irak die
2	27,90	die bosnischen Serben	37	–59,77	und im Irak
3	39,52	und der Opposition	38	–40,82	gegen den Irak
4	35,56	Aung San Suu	39	–27,7	amerikanischen im Irak
5	30,3	San Suu Kyi	40	–52,48	der den USA
6	32,93	der der Menschenrechte	41	–34,99	der USA im
7	31,71	der in Bosnien	42	–48,11	Abzug aus dem
8	30,3	der ersten Runde	43	–43,74	aus dem Irak
9	28,98	in der Serben	44	–41,82	gegen den Terrorismus
10	27,66	vor den Wahlen	45	–24,49	Kampf gegen den
11	27,66	die immer wieder	46	–21,87	Kampf gegen Terrorismus
12	26,34	die in gegen	47	–21,87	Krieg gegen den
13	26,34	Der Präsident der	48	–37,91	den neunziger Jahren
14	26,34	in der Slowakei	49	–35,28	hiess es in
15	26,34	am abend der	50	–33,53	aus dem Gazastreifen
16	25,36	der Nato und	51	–30,62	der in Kosovo
17	25,03	am in Paris	52	–29,16	Präsident Bush in
18	23,71	die absolute Mehrheit	53	–26,24	den besetzten Gebieten
19	23,71	russische Präsident Jelzin	54	–24,78	auf in Bagdad
20	22,39	der in Sarajewo	55	–24,78	die Kontrolle über
21	22,39	die Koalition der	56	–24,78	im Weissen Haus
22	22,39	der frühere der	57	–24,78	der Internationalen Atom-
23	22,39	sich die Frage			energieagentur
24	21,07	Beginn der in	58	–23,33	der Ministerpräsident Sharon
25	21,07	in den Reihen	59	–21,87	Usama bin Ladin
26	19,76	die Fortsetzung der	60	–20,3	der Suche nach
27	17,12	Kim Dae Jung	61	–18,95	die in Darfur
28	15,81	Kim Young Sam	62	–18,95	Debatte über die
29	17,12	die Wege geleitet	63	–17,49	in den Tod
30	15,81	gegen die Korruption	64	–16,48	eine Lösung des
31	15,81	der in Grosny	65	–16,04	die Opfer der
32	15,81	in der DDR	66	–16,04	Saddam Hussein der
33	15,81	Bundeskanzler Kohl in	67	–16,04	Zahl der auf
34	15,81	Präsident Clinton und	68	–16,04	der gemässigten Islamisten
35	15,65	die Tatsache dass	69	–15,3	hat den letzten

**Tabelle 13.1:** Typische Mehrworteinheiten im Ressort ‚Ausland‘ in den Zeitperioden 1995–1997 (links) und 2003–2005 (rechts). Manuelle Auswahl für die weitere Analyse (ausführlichere Variante im Anhang A.1, vollständige Version: <http://www.bubenhofer.com/korpusanalyse/>).

*nahmt, (Saddam) der Terrorisierung, Anschlag auf Saddam-Gericht Bagdad etc.*<sup>3</sup>

Ein weiterer Themenkreis lässt sich über die Mehrworteinheiten *Abzug aus dem* (42) *Gazastreifen* (50), *den besetzten Gebieten* (53) definieren, bei dem *Ministerpräsident Sharon* (58) eine Rolle spielt.

Auch die Mehrworteinheiten der Periode 1995–1997 (ebenfalls Tabelle 13.1 auf der vorherigen Seite) widerspiegeln Ereignisse wie die Jugoslawienkriege: *der/die bosnische Serben* (1f.), *der in Bosnien* (7), *in der Serben* (9), *der Nato und* (16) und *der in Sarajewo* (20) sind Indikatoren dafür. Die Friedensnobelpreisträgerin und Oppositionspolitikerin *Aung San Suu Kyi* (4) in Burma war in dieser Periode immer wieder Thema, ebenso der südkoreanische Oppositionspolitiker *Kim Dae Jung* (27) und der damalige *Bundeskanzler Kohl* (33), *US-Präsident Clinton* (34) und der *russische Präsident Jelzin* (19). Als Kollokatoren zu diesen Ausdrücken führt das Typikprofil für *Aung San Suu Kyi* über 120 unterschiedliche Mehrworteinheiten auf: *Opposition, Burma, praktisch/wieder unter Hausarrest, Freilassung, Rangun, Abriegelung*. Bei *Bundeskanzler Kohl* sind dies z. B. Kollokatoren wie *Besuch, und Scharping, der Bundeswehr, der Streitkräfte, und tschechische*. Bei *Präsident Jelzin: Rede ... Lage, in Tschetschenien, Moskauer ... Clinton, ihren Rücktritt, Kabinettsmitglieder*.

Neben den oben genannten Mehrworteinheiten, deren Existenz quasi historisch erklärt werden kann, gibt es andere, bei denen nicht auf den ersten Blick verständlich ist, weshalb sie für die eine oder andere Periode typisch sind. Dazu gehören für die erste Periode (1995–1997) *die immer wieder* (11), *die in gegen* (12), *sich die Frage* (23) usw., bzw. *hiess es in* (49), *Debatte über die* (62) oder *hat den letzten* (69) in der zweiten Periode (2003–2005). Ohne detailliertere Analysen lassen sich aufgrund dieser Mehrworteinheiten keine Hypothesen bilden. Der Ausdruck *Debatte über die*, der besonders signifikant für den Zeitraum 2003–2005 ist, könnte beispielsweise zur Vermutung führen, dass das Lexem *Debatte* in neuerer Zeit generell häufiger verwendet wird. Dem ist aber nicht so: Die Verteilung über die elf Jahre ist relativ

<sup>3</sup> Mehrworteinheiten, die nicht in den zusammenfassenden Tabellen in der vorliegenden Arbeit aufgeführt sind, können in den entsprechenden Typikprofilen online nachgeschlagen werden: <http://www.bubenhofer.com/korpusanalyse/>.

stabil.<sup>4</sup> Allerdings gibt es 68 unterschiedliche Mehrworteinheiten mit dem Lexem *Debatte*, die signifikant für die jüngere Zeitperiode sind, während es nur 15 sind, die für die ältere Periode signifikant sind. Das verwendete lexikalische Material in diesen Mehrworteinheiten ist in jüngerer Zeit also vielfältiger, als es in der älteren Periode der Fall ist.<sup>5</sup> Weitere Analysen müssten zeigen, wo der Grund für diese gestiegene Variabilität in den Sprechweisen im Zusammenhang mit Debatten liegt und ob sich das in anderen Korpora bestätigt. Ich beschränke mich im Folgenden aber auf drei andere tiefergehende Analysen für das Auslandsressort.

### 13.1.2 Die *bosnischen Serben*: Die Verwendung von Ethnienbezeichnungen

Die Verteilung der Artikel, in denen der Ausdruck *bosnische(n) Serben* vorkommt, ergibt ein deutliches Bild einer variablen Typik, wie Tabelle 13.2 auf der nächsten Seite zeigt. Bereits oben in der Grobanalyse habe ich dargestellt, dass das Schwergewicht der Nennungen von *bosnischen Serben* 1995 und 1996 mit dem Bosnienkrieg erklärt werden kann, der zu dieser Zeit stattfand. Allerdings muss weiter gefragt werden, ob diese Mehrworteinheit zu einem Muster ETHNIENBEZEICHNUNG abstrahiert werden kann, das bestimmte Funktionen in Diskursen einnimmt.

Um der Frage nach den Ethnienbezeichnungen nachzugehen, bietet es sich an, die paradigmatischen Konstellationen (vgl. Kapitel 8.2.3) von ETHNIENBEZEICHNUNGEN zu untersuchen. Dies zielt darauf, herauszufinden, welche alternativen Ausdrucksweisen es für ETHNIENBEZEICHNUNGEN in Kontexten gibt, die für solche Ausdrucksweisen theoretisch offen stehen. Als alternative Ausdrucksweise sind Bezeich-

<sup>4</sup> Durchschnittlich 5,3 % aller Artikel im Ressort ‚Ausland‘ im Zeitraum 1995–2005 enthalten das Lexem *Debatte* und es gilt:  $\chi^2 = 4,488$ ,  $df = 10$ ,  $p > 0,10$  (nicht signifikant).

<sup>5</sup> In den für die Zeit von 1995–1997 signifikanten Mehrworteinheiten mit dem Bestandteil *Debatte* werden folgende Lemmata verwendet: *Debatte, Debatten, Kammer, Kolyma, Massnahmen, Misstrauensdebatte, Thema, der, die, gegen, in, um, und*.

In der Zeit von 2003–2005 sind es die folgenden Lemmata: *Ansätze, Bestätigung, Buffets, Debatte, EU, Frankreichs, Geschichte, Kapitel, Kommunisten, KPF, Länder, Provinz, Reformdebatte, Türkei, Uno-Debatte, Verfassung, Waffenexporte, Zukunft, an, chaotische, der, die, dunkle, eine, einer, europäische, fordern, fünf, in, öffentliche, polnische, über, um, und, unter, zu*.

Jahr	<i>bosnische(n) Serben</i> #	%	Total #
1995	32	6,36	503
1996	10	2,15	466
1997	5	1,07	469
1998	3	0,64	470
1999	2	0,38	529
2000	0	0	567
2001	4	0,72	552
2002	3	0,58	519
2003	4	0,69	576
2004	0	0	515
2005	2	0,4	504
Total	65		5670

$\chi^2 = 147,122$ ,  $df = 10$   
 $p < 0,001$  (signifikant),  $V = 0,161$   
 $\phi = 5,9091 \pm 8,6493$

**Tabelle 13.2:** Verteilung und  $\chi^2$ -Statistik (Einheit: Artikel) von *bosnische(n) Serben* im NZZ-Korpus.

nungen wie *Einwohner/innen*, *Bürger/innen* oder Kombinationen mit Staats- oder Länderbezeichnungen denkbar. In einem Artikel mit dem Thema ‚Bosnien/Serbien‘ könnte statt von *bosnischen Serben* auch von *Bewohnern (Bosniens)* und/oder *Menschen* die Rede sein. Oder das Gewicht könnte generell weniger stark auf Ethnienbezeichnungen liegen, wie z. B. in Ausdrucksweisen wie *die Wirtschaft Serbiens steckt in der Krise* statt *die Serben stecken wirtschaftlich in der Krise*. Wann wird welche Ausdrucksweise bevorzugt?

Die Jugoslawienkriege sind vielschichtig und langandauernd. Da als Ausgangspunkt die Mehrwortereinheit *bosnische Serben* dient, beschränke ich mich auf die Regionen Kroatien, Bosnien und Serbien, und auf das Ende und die Folgen der Kriege in diesen Regionen. So arbeite ich mit einer Auswahl von Artikeln, in denen mindestens eine Länder- oder Ethnienbezeichnung dieser Regionen vorkommt. Diese Auswahl stellt ein Teilkorpus für die weiteren Analysen dar: Darin werden die Anteile von Ethnien- und Länderbezeichnungen berechnet. Für die Erstellung des Teilkorpus wurde eine Reihe von Suchbegriffen definiert, die als konstitutiv für das Thema ‚Kroatien/Bosnien/Serbien‘ angenommen werden können. Mit Hilfe dieser Suchbegriffe wurden

die Artikel für das Teilkorpus des NZZ-Korpus bestimmt.<sup>6</sup> Dieses Teilkorpus umfasst 555 Artikel in 621 Zeitungsausgaben; die Artikel umfassen 505 419 Wörter. In diesem Teilkorpus wurden wiederum die Anteile von Artikeln bemessen, in denen eine Reihe von Bezeichnungen für Ethnien und Länder/Staaten der Region genannt werden.

In Abbildung 13.1 auf der nächsten Seite sind die Anteile der Artikel des Teilkorpus dargestellt, die mindestens eine der Ethnien *Bosnier/Bosnjaken*, *Kroaten*, *Serben* (oder ihre weiblichen Formen)<sup>7</sup> enthalten. Ebenso sind die Anteile der Artikel mit entsprechenden Länderbezeichnungen angegeben. Die Befunde sind in Abbildung 13.2 auf Seite 216 zusammengefasst. Es werden jeweils die Anteile der beiden Kategorien ETHNIE und LÄNDERBEZEICHNUNG in ihrem Verhältnis zueinander ausgedrückt.

Welche Beobachtungen lassen sich aufgrund dieser Daten machen?

1. Unter der Menge der Artikel mit dem Thema ‚Bosnien/Kroatien/Serbien‘ gehen mit dem Ende des Bosnien- und Kroatienkrieges 1995 die Nennungen von Ethnienbezeichnungen zurück.
2. Im selben Zeitraum geht auch die Bezeichnung *Bosnien* bzw. *Herzegowina* zurück.
3. Die Länderbezeichnungen *Kroatien*, *Serbien* und *Jugoslawien* nehmen mit der Abnahme der Ethnienbezeichnungen zu.

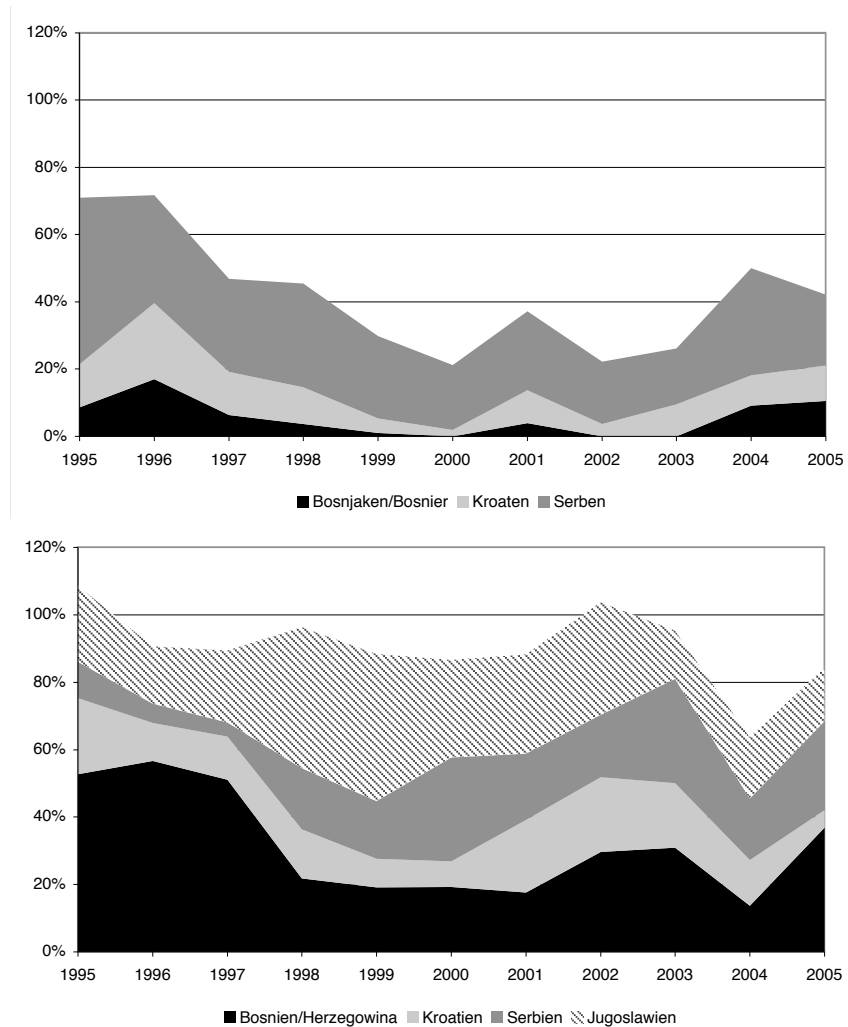
Folgende möglichen Hypothesen lassen sich daraus ableiten: a) Frequente Verwendungen von Ethnienbezeichnungen in Zeitungsartikeln sind Indikatoren für gesellschaftliche, z. B. kriegerische, Auseinandersetzungen. b) Die frequente Verwendung von Länderbezeichnungen dagegen zeigen eher eine Normalisierung der Lage an.

Die folgenden Belege mögen dies illustrieren:

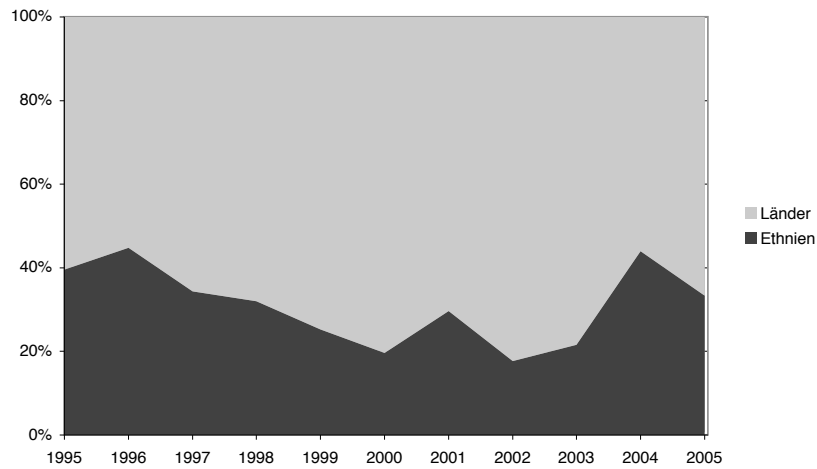
- (19) Die **bosnischen Serben** haben nach Angaben der Uno in Sarajewo die Blockade eines Schützenpanzers mit sechs französischen Blauhelm-Soldaten an Bord eingestellt. Die **Serben**

<sup>6</sup> Es handelt sich dabei um Artikel des Auslandressorts, in denen mindestens einer der folgenden Wortteile vorkommen: *jugoslaw*, *serb*, *bosn*, *kroat*.

<sup>7</sup> Allerdings gibt es in den untersuchten Daten kaum weibliche Formen dieser Bezeichnungen.



**Abbildung 13.1:** Paradigmatische Konstellation für ETHNIENBEZEICHNUNGEN (oben) und LÄNDERBEZEICHNUNGEN (unten) in Relation zu Artikeln mit dem Kontext ‚Kroatien/Bosnien/Serbien‘. Anzahl Artikel des Teilkorpus: 555; Ethnienbezeichnungen: Nur Pluralformen, auch weibliche Formen.



**Abbildung 13.2:** Paradigmatische Konstellation für ETHNIENBEZEICHNUNGEN und LÄNDERBEZEICHNUNGEN in Relation zu Artikeln mit dem Kontext ‚Kroatien/Bosnien/Serbien‘ (zusammengefasst). Anzahl Artikel des Teilkorpus: 555.

hätten den **Franzosen** erlaubt, den Stützpunkt Sierra Vier zu verlassen und zu ihrer Basis am Flughafen Sarajewo zu fahren, nachdem sie in ihrem Panzerwagen über fünf Tage blockiert waren. Der Uno-Sprecher wertete den Vorgang als mögliches Zeichen dafür, dass die **Serben** von ihrem Konfrontationskurs gegenüber der Uno abrücken könnten. *Neue Zürcher Zeitung vom 31. Mai 1995, Ressort ‚Ausland‘, Cyrill Stieger: „Kriegerische Töne Karadzics. Aufkündigung aller Vereinbarungen mit der Uno“.*

- (20) Aus den von den **Serben** gehaltenen Gebieten im Nordwesten Bosniens um die Stadt Banja Luka sind laut Angaben des IKRK seit August 20 000 **Kroaten** und **Bosnjaken** vertrieben worden. *Neue Zürcher Zeitung vom 13. Oktober 1995, Ressort ‚Ausland‘, Reuters: „Beginn der Waffenruhe in Bosnien-Herzegowina. Stolpersteine auf dem langen Weg zum Frieden“.*

- (21) Nach Meinung von Zoran Djindjic, dem Präsidenten der Demokratischen Partei, hat Milosevic mit dem Abbruch der politischen Beziehungen zu Karadzic seinen bisher grössten Fehler begangen. Es sei ihm nicht gelungen, seinen gefährlichsten Rivalen zu stürzen, der sich als Führer jener **Serben** versteht, die westlich der Drina leben. *Neue Zürcher Zeitung vom 16. Februar 1995, Ressort ‚Ausland‘: „Serbiens zerstrittene Opposition. Neue Fronten nach dem Bruch mit Karadzic“.*

Die Belege oben zeigen die Verwendung von Ethnienbezeichnungen im kriegerischen Kontext, wobei auffällt, dass selbst die französischen Uno-Soldaten als *Franzosen* bezeichnet sind. Die Zugehörigkeit zu einer Ethnie oder Nation scheint im kriegerischen Kontext eine wichtige Kategorie zu sein.

Natürlich finden sich auch in nicht-kriegerischen Kontexten Ethnienbezeichnungen, allerdings in viel geringerer Zahl. Ein Beispiel ist eine Nachricht zum Bericht der Chefanklägerin des Uno-Tribunals, Carla del Ponte, die Kroatien vollständige Kooperation mit dem Gericht bescheinigt:

- (22) Tatsächlich hat Del Ponte anlässlich ihres Besuchs in Zagreb am letzten Freitag den **Kroaten** wenig Anlass gegeben, mit einem für sie derart positiven Bericht rechnen zu können. *Neue Zürcher Zeitung vom 4. Oktober 2005, Ressort ‚Ausland‘, Woker M.: „Freude in Zagreb über Del Pontes Entscheid“.*

Die Frage bleibt, ob die vorliegenden Befunde generalisiert werden können. Einen Hinweis dazu gibt eine Recherche zum Vorkommen von *den/die Deutschen* im NZZ-Korpus.<sup>8</sup> Die Verteilung über die Jahre 1995–2005 ist nicht signifikant, wie Tabelle 13.3 auf der nächsten Seite zeigt.

Aufschlussreich ist aber ein genauerer Blick auf die Kontexte. So handelt es sich bei 35 % der Artikel um Sportberichterstattung (vgl. Tabelle 13.4 auf der nächsten Seite). Unter den restlichen Fällen (157 Nennungen in 135 Artikeln) können 57 Nennungen dem Kontext ‚Zweiter Weltkrieg‘ zugeordnet werden. Das bedeutet, dass mit *die*

<sup>8</sup> Die Abfrage erfolgte mit dem regulären Ausdruck `(die|den)\nDeutschen\n[[:lower:]]+`, um nur Pluralformen der Ethnienbezeichnung zu erhalten und damit adjektivische Verwendungen ausschließen zu können.



Jahr	<i>die/den Deutschen</i>		Total #
	#	%	
1995	24	0,65	3666
1996	23	0,62	3711
1997	20	0,52	3827
1998	22	0,61	3616
1999	21	0,49	4248
2000	17	0,38	4496
2001	17	0,39	4373
2002	15	0,34	4385
2003	19	0,45	4235
2004	15	0,36	4155
2005	16	0,39	4131
Total	209		44 843

$\chi^2 = 10,953$ , df = 10  
 $p > 0,10$  (nicht signifikant),  $V = 0,016$   
 $\odot = 19 \pm 3,07$

**Tabelle 13.3:** Verteilung und  $\chi^2$ -Statistik (Einheit: Artikel) von *die/den Deutschen* im NZZ-Korpus.

Ressort	<i>die/den Deutschen</i>		<i>deutsch</i>		<i>Deutschland</i>	
	#	%	#	%	#	%
Sport	74	35	474	9	686	17
Feuilleton	45	22	1260	24	478	12
Ausland	27	13	583	11	472	12
Wirtschaft	14	7	1192	23	1280	31
Leserbriefe	8	4	93	2	71	2
Inland	7	3	348	7	281	7
andere	34	16	1290	25	827	20
Total	209	100	5240	100	4095	100

**Tabelle 13.4:** Die Vorkommen von *die/den Deutschen*, *deutsch* und *Deutschland* in den Ressorts des NZZ-Korpus (absolute und relative Frequenzen).

*Deutschen* Soldaten oder Zivilpersonen aus der Zeit Nazideutschlands gemeint sind, wie folgender Beleg zeigt:

- (23) Die historische Wirklichkeit besteht ganz kurz gesagt darin, dass **die Deutschen** während des Kriegs aus der besetzten Ukraine Männer und Frauen auch nach Singen brachten, wo sie unter erbärmlichsten Bedingungen Zwangsarbeit leisten mussten. *Neue Zürcher Zeitung* vom 25. Januar 1997, Ressort ‚Feuilleton‘, Egger, C.: „Solothurn, fast wie zu alten Zeiten. Stoff zu Diskussionen an den Filmtagen“.

Neben dieser größten Gruppe verteilen sich die verbleibenden 100 Nennungen auf unterschiedliche Arten von Verwendungsweisen. Die Mehrzahl der Kontexte sind beschreibend und fassen *die Deutschen* als Volk, das (oft im Vergleich mit anderen Nationen) eine bestimmte Eigenschaft aufweist:

- (24) Ein europäischer Vergleich ergab, dass **die Deutschen** mit Abstand das höchste Porto bezahlen. Die Post hat nun zwei Monate Zeit, um die detailliert aufgeführten wettbewerblichen Bedenken der Kommission zu zerstreuen. *Neue Zürcher Zeitung* vom 9. August 2000, Ressort ‚Wirtschaft‘, Lautenschütz R.: „Drittes EU-Verfahren gegen die Deutsche Post. Häufung von Wettbewerbsklagen gegen deutsche Firmen“.

Oder der Ausdruck wird in politischen Kontexten verwendet, bei denen (aus schweizerischer Perspektive) auf deutschländische Vorgänge Bezug genommen wird, oder in internationaler Politik mehrere Nationen verglichen werden:

- (25) Allerdings würde es wohl genau auf eine Fortsetzung dieser Praxis hinauslaufen, wenn **die Deutschen** im September bei Neuwahlen eine „Klärung der Verhältnisse“ im Sinne Schröders und Münteferings herbeiführten. *Neue Zürcher Zeitung* vom 28. Mai 2005, Ressort ‚Ausland‘, Dedial J.: „Im Schatten des Scheiterns“.
- (26) Vor allem vermied Schröder es, auch nur anzudeuten, was er zu tun beabsichtigt, um sich und **die Deutschen** den Amerikanern wieder als Partner und Freunde in Erinnerung zu rufen.

*Neue Zürcher Zeitung vom 4. April 2003, Ressort ‚Ausland‘, Dedial J.: „Aus Berlin nicht viel Neues“.*

Weitere Belege thematisieren nationale Eigenheiten oder Charakterisierungen:

- (27) Aber für die Schweizer gilt gleichfalls das, was Bundespräsident Herzog über **die Deutschen** sagte, nämlich: Wir sind nichts Besonderes, aber etwas Bestimmtes. *Neue Zürcher Zeitung vom 11. März 1995, Ressort ‚Zeitfragen‘: „Eine paradoxe Forderung: Spezifische Identität bekräftigen“.*
- (28) Unweigerlich taucht die Frage auf, ob **die Deutschen** denn ein Volk von „Schnäppchenjägern“ seien. Jüngere Umfragen unter dem Publikum deuten in der Tat darauf hin, dass die Leute heute schärfer auf die Preise schauen als früher. *Neue Zürcher Zeitung vom 2. März 1996, Ressort ‚Fokus der Wirtschaft‘, Gygi B.: „Ist der Lebensmitteleinzelhandel eine wohltätige Einrichtung? Symbiose von ‚Schnäppchenjägern‘ und Discountern in Deutschland“.*

Die Beobachtungen zeigen, dass die Bezeichnung *die Deutschen* im NZZ-Korpus stark mit dem Kontext ‚Zweiter Weltkrieg‘ korreliert. Es gibt nicht viele andere Kontexte, in denen es nötig scheint, von *den Deutschen* statt von *Deutschland* oder *deutsch* zu sprechen; die beiden letzteren Bezeichnungen kommen im Korpus denn auch sehr viel häufiger vor (vgl. Tabelle 13.4 auf Seite 218).

### 13.1.3 Kontextualisierungsprofile: *Kampf gegen X, Kampf dem X oder Kampf mit X?*

Die berechneten Mehrworteinheiten *Kampf gegen Terrorismus*, *im Kampf gegen* etc. aus Tabelle 13.1 auf Seite 210 motivieren, die Kontextualisierungsprofile (vgl. Kapitel 8.2.2) dieser Ausdrücke genauer zu analysieren. Es scheint zwar plausibel, dass *Kampf gegen Terror*<sup>9</sup> ein Ausdruck ist, der typisch für die Zeit nach dem 11. September

<sup>9</sup> Der Stern (\*) steht für beliebig viele zusätzliche Zeichen (außer Leerzeichen). Hier und im Folgenden wird damit angezeigt, dass in der Suche auch andere Endungen des Wortes gefunden würden. *Terror\** findet *Terror*, *Terrorismus*, *Terroristen* etc.

Jahr	<i>Kampf(s/es)</i> gegen ... <i>Terror*</i>		<i>Kampf(s/es)</i> gegen		Total
	#	%	#	%	#
1995	2	0,05	62	1,69	3666
1996	3	0,08	58	1,56	3711
1997	1	0,03	55	1,44	3827
1998	3	0,08	58	1,60	3616
1999	5	0,12	44	1,04	4248
2000	2	0,04	61	1,36	4496
2001	19	0,43	82	1,88	4373
2002	22	0,50	67	1,53	4385
2003	16	0,38	70	1,65	4235
2004	17	0,41	71	1,71	4155
2005	18	0,44	65	1,57	4131
Total	108		693		44 843
$\chi^2 = 64,222$		$\chi^2 = 13,403$			
$p < 0,001$ (signifikant)		$p > 0,10$ (nicht sign.)			
$V = 0,038$		$V = 0,017$			
$\phi = 9,8 \pm 8$		$\phi = 63 \pm 9,4$			

**Tabelle 13.5:** Verteilungen und  $\chi^2$ -Statistiken (Einheit: Artikel) für die Mehrworteinheiten *Kampf(s/es) gegen ... Terror\** bzw. *Kampf(s/es) gegen* im NZZ-Korpus (alle Ressorts; für die  $\chi^2$ -Statistiken gilt:  $df = 10$ ).

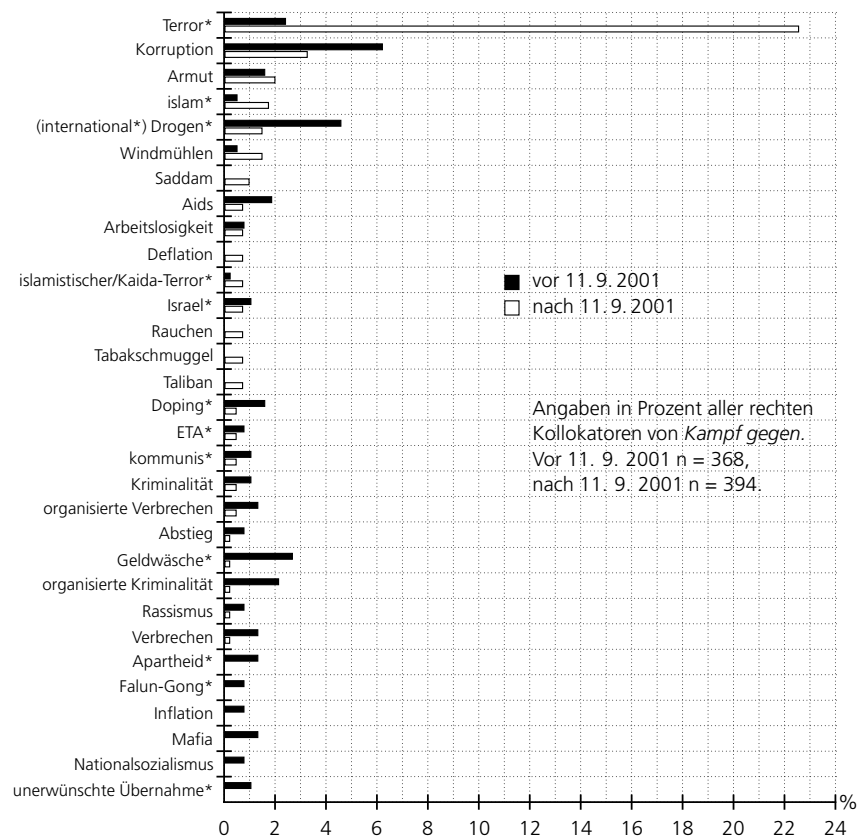
2001 ist (variable Typik), wie Tabelle 13.5 zeigt. Doch allgemeiner ist zu fragen nach dem Gebrauch des Musters KAMPF GEGEN X und den Füllungen des ‚Slots‘ X.

Tabelle 13.5 zeigt die Frequenzen von *Kampf gegen* im NZZ-Korpus. Die Verteilung ist nicht signifikant, die Verwendung also über die Jahre ungefähr ähnlich (stabile Typik). Abbildung 13.3 auf der nächsten Seite zeigt die Füllungen von X im Muster KAMPF GEGEN X für die Perioden vor und nach dem 11. September 2001.<sup>10</sup>

Für die weitere Analyse interessiert im Besonderen, ob und wie Diskurse die Sprechweise verändern. Im Beispiel von KAMPF GEGEN X wäre also zu fragen:

1. Gibt es zu bestimmten Zeiten oder in bestimmten Kontexten typische Füllungen des Slots X im Muster KAMPF GEGEN X?

<sup>10</sup> Ich greife hier die in Kapitel 8.2.2 thematisierte Abbildung 8.5 auf Seite 162 wieder auf. Zu den Details der Berechnung vgl. die Ausführungen ebendort.



**Abbildung 13.3:** Die Füllungen von X im Muster KAMPF GEGEN [ARTIKEL] X mit Mindestfrequenz 3 für die Zeit vor und nach dem 11. September 2001. ,\*‘ steht als Platzhalter für weitere Zeichen. Details zur Berechnung können dem Text entnommen werden.

2. Welche alternativen Ausdrucksmöglichkeiten zum Muster KAMPF GEGEN X existieren? Und wann werden sie verwendet?

Die notwendigen Analysen, um die erste Fragestellung zu verfolgen, sind einsichtig: In erster Linie geht es darum, die Füllung von X im Muster KAMPF GEGEN X in Abhängigkeit von Zeit, Thema, Textsorte etc. zu untersuchen. Die Analysen um die zweite Fragestellung zu verfolgen, können in folgende Richtung gehen:

Als engere alternative Ausdrucksmöglichkeiten von KAMPF GEGEN X liegen folgende Varianten auf der Hand: KAMPF MIT X und KAMPF DEM/DER [DATIVOBJEKT].<sup>11</sup> Es ist zu fragen, ob der propositionale Gehalt der drei Varianten gleich ist. Die möglichen syntaktischen Einbettungen der Ausdrücke zeigt, dass die Ausdrucksvarianten erst als ausgebauter Satzteil syntaktisch austauschbar werden, da sie teilweise syntaktische Restriktionen aufweisen:

1. *Kampf gegen den Terrorismus*
  - a) *Jemand fordert den Kampf gegen den Terrorismus.*
  - b) *Jemand befindet sich im Kampf gegen den Terrorismus.*
2. *Kampf mit dem Terrorismus*
  - a) *Jemand fordert den Kampf mit dem Terrorismus.*
  - b) *Jemand befindet sich im Kampf mit dem Terrorismus.*
3. *Kampf dem Terrorismus*
  - a) *Jemand fordert: Kampf dem Terrorismus!*
  - b) *Heute gilt: Kampf dem Terrorismus!*

Die (erfundenen) Beispielsätze lassen vermuten, dass die Ausdrucksvarianten nicht alle dieselbe Konnotation aufweisen müssen und dass vor allem der Ausdruck *Kampf dem Terrorismus* eine ganz bestimmte pragmatische Funktion aufweist, nämlich als Parole wirkt.

<sup>11</sup> Weitere Ausdrucksvarianten könnten mit dem Lexem *Krieg* gebildet werden. Ich sehe der Einfachheit halber davon ab. Die Ausdrucksvariante *Krieg gegen den Terror* wurde schon mehrfach öffentlich kritisiert und ruft sogar juristische Kritik hervor: „Acht Juristen haben in Genf einen im Auftrag der International Commission of Jurists verfassten Bericht vorgestellt. Darin wird die Aufhebung des Begriffs ‚Krieg gegen den Terror‘ gefordert, der vielen Verstößen gegen die Menschenrechte Vorschub geleistet habe.“ *Neue Zürcher Zeitung vom 17 Februar 2009, Ressort ‚Ausland‘*.

KAMPF MIT X [ARTIKEL]	KAMPF DEM/DER [DATIVOBJEKT]
allen Mitteln gegen Israel; Aussicht auf Erfolg; baskischen Separatisten und Terroristen; Bayern München; Behörden; Berg; Depression; Drachen; drohenden Überschuldung; eigenen Schwächen; Elementen; erbarmungslosen Feind; Gegnern; gleich langen Spiessen; gleicher Waffe; Grammatik; Great-West Lifeco; harten Bandagen; HIV; ihnen; ihren Texten; Immunsystem; Klavierschülergeschlecht; Konkurrenz; Leuten; Mächten einer niederdrückenden Ideologie; nassen Element; Neat-Gegnern; Schönheit; sehr jungen Gegnern; sich selber; sich selbst; Trieb; ungewissem Ausgang; vorpreschenden Spunden; Wettbewerbskommission; Wind, Wasser und Wellen; x Toren; Zeit	„Bussen- und Gebührenterror“; Dengue-Fieber; Diskriminierung von HIV-Positiven; Feinstaubbelastung; Guineawurm; Korruption; Kriminalität; Lobbying; Minenplage; Neoliberalismus; Schuldenmisswirtschaft; Schwefel; Spam; Stau; Strafflosigkeit

**Tabelle 13.6:** Die Füllungen von X in den Mustern KAMPF MIT X und KAMPF DEM/DER [DATIVOBJEKT] im NZZ-Korpus. Bis auf *Kampf mit dem Berg* (vier Vorkommen) gibt es für jede Mehrworteinheit jeweils nur einen Beleg im Korpus.

Das NZZ-Korpus enthält keine Belege für die Varianten 2 und 3, und nur wenige Belege für die von Varianten 2 und 3 abstrahierten Muster KAMPF MIT X und KAMPF DEM/DER [DATIVOBJEKT]. Die Füllungen für X in den beiden Mustern sind in Tabelle 13.6 aufgeführt. Dabei wird sofort klar, dass sich in den untersuchten Daten deutliche Unterschiede im Sprachgebrauch zeigen, wobei es aber schwierig ist, die Unterschiede zu benennen. Vielleicht lässt es sich so fassen: KAMPF DEM/DER [DATIVOBJEKT] wird überhaupt nicht, und KAMPF MIT X [ARTIKEL] nur vereinzelt in kriegesischen Kontexten verwendet. Etwas anders zeigt sich das Bild bei KAMPF GEGEN X (vgl. Abbildung 13.3 auf Seite 222). Bei diesem Muster fallen die hochfrequenten Kollokatoren wie *Terror\**, *Saddam*, *islamistischer/Kaida-Terror\**, *Taliban*, *ETA* etc. auf, die in kriegesischen Kontexten zu finden sind.

Dieser Befund steht aufgrund der Datenlage auf schwachen Füßen. Es gibt im NZZ-Korpus zu wenig Belege für die beiden Muster KAMPF MIT X und KAMPF DEM/DER [DATIVOBJEKT]. Es bietet sich deshalb an, ein weiteres Korpus, beispielsweise das DeReKo IDS (o.J.)-Korpus des Instituts für Deutsche Sprache, als Referenzkorpus heranzuziehen.

KAMPF(S/ES) GEGEN X	#	%	KAMPF(S/ES) MIT X	#	%
Abstieg	453	4,53	[Sportresultat X: Y]	344	11,21
Terror(ismus)	329	3,29	Behörde(n)	32	1,04
organisiert*	225	2,25	sich selbst/sich selber	31	1,01
Doping	183	1,83	Drachen	26	0,85
Korruption	161	1,61	Natur	21	0,68
Armut	129	1,29	Bürokratie(n)	20	0,65
Drogen	119	1,19	Tücke(n) des/der	17	0,55
Geldwäsche/			allen Mitteln	16	0,52
Geldwäscherei	105	1,05	Elementen	16	0,52
Kriminalität	96	0,96	Konkurrenten/		
Uhr	94	0,94	Konkurrenz	16	0,52
Mafia	90	0,90	Waffe(n)	16	0,52
AIDS	79	0,79			
illegale(n)	77	0,77			
Windmühle(n)	77	0,77			
internationale(n)	76	0,76			
Krebs	62	0,62			
seine(n)	61	0,61			
Rassismus	58	0,58			
Israel	51	0,51			
Verbrechen	51	0,51			
Total	10 000	100,00	Total	3070	100,00

**Tabelle 13.7:** Die Füllungen von X in den Mustern KAMPF GEGEN X und KAMPF MIT X im IDS-Korpus, deren Frequenzen mindestens 0,5 % aller Kollokatoren ausmachen.

Grundlage für die Recherche waren alle im IDS-Korpus öffentlich verfügbaren Tageszeitungen aus Deutschland, Österreich und der Schweiz im Zeitraum von 1991 bis 2006.<sup>12</sup> Die Tabellen 13.7 und 13.8 auf der nächsten Seite zeigen die häufigsten Füllungen für X in den drei Mustern.<sup>13</sup>

Mit Hilfe dieser Analysen lässt sich der oben formulierte Befund präzisieren:

<sup>12</sup> Das so zusammengestellte Korpus enthält damit 4 129 847 Texte mit insgesamt 960 395 973 Wörtern. Folgende Zeitungen sind in unterschiedlichen Zeiträumen vertreten: Berliner Morgenpost, Die Presse, Frankfurter Rundschau, Hamburger Morgenpost, Kleine Zeitung, Mannheimer Morgen, Neue Kronen-Zeitung, Oberösterreichische Nachrichten, Salzburger Nachrichten, St. Galler Tagblatt, Tiroler Tageszeitung, Vorarlberger Nachrichten, Zürcher Tagesanzeiger.

<sup>13</sup> Für das Muster KAMPF DEM/DER [DATIVOBJEKT] mussten im IDS-Korpus Belege mit Dativobjekten mit femininem Genus (*Kampf der Arbeitslosigkeit*) ignoriert werden, da das Korpus nicht syntaktisch annotiert ist und deshalb Dativobjekte nicht von Genitivobjekten (*der Kampf der Regierung*) unterschieden werden konnten.



KAMPF DEM X	#	%	(Fortsetzung...)	#	%
Krebs	33	5,45	Faschismus	4	0,66
Stau	32	5,28	Hochwasser	4	0,66
Hunger	21	3,47	Müll	4	0,66
Dickdarmkrebs	13	2,15	Schimmel*	4	0,66
Brustkrebs	10	1,65	Tod	4	0,66
Elektrosmog	9	1,49	Übergewicht	4	0,66
Fahrraddiebstahl	9	1,49	Atomtod	3	0,50
Herztod	8	1,32	blauen Dunst	3	0,50
Schlaganfall	8	1,32	Chaos	3	0,50
Terror*	8	1,32	Darmkrebs	3	0,50
Feuerbrand	7	1,16	Feinstaub	3	0,50
Rassismus	7	1,16	Grünen Star	3	0,50
Sterilen und Leblosen	7	1,16	Hanf	3	0,50
Alkohol	6	0,99	Kommunismus	3	0,50
Kinderkrebs	6	0,99	Mißbrauch	3	0,50
Pusch	6	0,99	Orgasmus	3	0,50
Doping	5	0,83	Proporz	3	0,50
Kindsmissbrauch	5	0,83	Schilderwald	3	0,50
Verbrechen	5	0,83	Sozialmissbrauch	3	0,50
Winterspeck	5	0,83	Verkehrschaos	3	0,50
Drogen*	4	0,66			
Total				606	100,00

**Tabelle 13.8:** Die Füllungen von X im Muster KAMPF DEM X im IDS-Korpus, deren Frequenzen mindestens 0,5 % aller Kollokatoren ausmachen.

1. KAMPF GEGEN X: Mit den dominanten Füllungen verweist der propositionale Gehalt des Ausdrucks auf eine internationale, politische Dimension, die wohl als ‚gesellschaftlich bedeutend‘ umrissen würde. Ausnahme davon: Der *Kampf gegen den Abstieg*, der im sportlichen Kontext verwendet wird.
2. KAMPF MIT X: Mit den dominanten Füllungen beschreiben die Ausdrücke eher Anekdotisches (*Kampf mit den Behörden*, der *Bürokratie*, den *Tücken [der Technik]*), oder referieren auf die eigene Persönlichkeit (*Kampf mit sich selbst*). Die Verwendung ist aber sehr heterogen. An der Spitze stehen auch hier Verwendungen im Zusammenhang mit Sportresultaten: *Den Kampf mit 2 : 0 (gewonnen)*.
3. KAMPF DEM X: Auffallend sind hier die dominanten Verwendungen mit dem Lexem *Krebs*- und Bezeichnungen für andere

Krankheiten. Mehrheitlich bewegen sich die Ausdrücke eher auf der individuellen und durchaus tragischen Ebene, daneben werden sie aber auch in Kontexten der Innenpolitik oder von Problemen im Bereich ‚Naturgewalten‘ und ‚Verkehr‘ genannt.

Neben diesen Präferenzen für Füllungen von X gibt es auch Belege, die zeigen, dass die Muster teilweise austauschbar sind. So finden sich sowohl Belege für *Kampf gegen das Verbrechen* als auch *Kampf dem Verbrechen* oder *Kampf gegen Rassismus* und *Kampf dem Rassismus* und andere mehr.

Interessant im Zusammenhang mit *Terror* ist besonders die seltene Verwendung von *Kampf dem Terror*\* im Vergleich zu anderen Füllungen des Ausdrucks KAMPF DEM X. Gibt es, neben den jeweiligen syntaktischen Restriktionen, einen semantischen oder pragmatisch-funktionalen Unterschied zwischen *Kampf dem Terror*\* und *Kampf gegen den Terror*? Fordert der politische Terror-Diskurs eine bestimmte Sprachgebrauchsvariante?

Alle acht Belege für *Kampf dem Terror*\* im IDS-Korpus sind Verwendungen in Titeln oder als Schlagzeilen von Zeitungsartikeln und haben Parolen-Charakter. Bei dreien können direkte oder indirekte Zitate dahinter vermutet werden:

- (29) Vranitzky, Schüssel: **Kampf dem Terror**, ob von links oder rechts! *Neue Kronen-Zeitung* vom 27. April 1995, S. 2.
- (30) US-Gegenoffensive mit Atomwaffen? „**Kampf dem Terrorismus** mit allen verfügbaren Mitteln“ *Salzburger Nachrichten* vom 24. August 1998, Ressort: ‚Weltpolitik‘.
- (31) **Kampf dem Terrorismus**. Der neue FBI-Chef Louis Freeh bezeichnete es als eine seiner wichtigsten Aufgaben, scharf gegen den Terrorismus vorzugehen. *Die Presse* vom 4. September 1993, Ressort ‚IN KÜRZE‘.

Bei übersetzten Zitaten können sprachliche Feinheiten, wie die Unterscheidung von *Kampf gegen* und *Kampf dem*, verloren gehen. Doch aus diskursanalytischer Sicht ist es trotzdem interessant zu sehen, welche Sprachgebrauchsvariante auch in der Übersetzung für einen bestimmten Zweck verwendet wird. Aufgrund der gemachten Beobachtungen könnte z. B. die These aufgestellt werden, dass mit *Kampf*

*dem Terror\** eine Sprachgebrauchsvariante verwendet wird, die politische Rhetorik, und vielleicht sogar Distanzierung des Verfassers/der Verfasserin gegenüber dieser Rhetorik, signalisieren soll. Weitere Evidenz für diese These ergibt eine unsystematische Suche in deutschsprachigen Tageszeitungen, wie folgende Belege zeigen:

- (32) Die konservative Zeitung fordert mehr **Kampf dem Terrorismus**: „Deutschland wird in den kommenden Jahren weiter massiv in den Ausbau der Sicherheitsarchitektur investieren müssen. Es wäre gut, wenn die staatstragenden Parteien, zu denen man neben Union und SPD auch Grüne und FDP zählen möchte, sich darüber gemeinsam Gedanken machen.“ *Sonn- tagsZeitung vom 9. September 2007, S. 23.*
- (33) Geiger ist sich sicher, dass für den Finanzplatz Schweiz die Bedeutung des Bankgeheimnisses steigen wird. Denn mit der Begründung „**Kampf dem Terrorismus**“ nähmen die Zugriffs- begehrllichkeiten auf die Privatsphäre in vielen Ländern zu. *Süddeutsche Zeitung vom 29. November 2006, S. V2/2.*
- (34) Ein weiteres Beispiel für Abzocke in großem Stil: die Ge- bührenerhöhung für den neuen Reisepass um 127 Prozent, natürlich unter dem Deckmantel „**Kampf dem Terrorismus**“. *Stuttgarter Zeitung vom 31. 10. 2005, Ressort ‚Leserbriefe‘, S. 40.*
- (35) Das Schlagwort „**Kampf dem Terrorismus**“ muss herhalten, die Folgen dieses Kriegs um Einfluss und Öl zu rechtfertigen. Sind nicht auch die Deutschen mitschuldig, wenn die ame- rikanischen Stützpunkte im Land für Militärtransporte etc. benützt werden? *Stuttgarter Zeitung vom 16. November 2004, Ressort ‚Leserbriefe‘, S. 8.*

Die Liste ließe sich fortsetzen. Es zeigt sich, dass die Verwendung von *Kampf dem Terror\** hier (z. B. in Leserbriefen) in argumentativer Funktion<sup>14</sup> verwendet wird, um diesen ‚Kampf‘ zu kritisieren.

Die bisherigen Analysen führen zu folgenden Hypothesen:

<sup>14</sup> Im Anschluss an die Argumentation in Kapitel 4.5 ist ein Textfragment dann ‚argu- mentativ‘, wenn es in die Struktur eines Toulmin-Schemas eingepasst werden kann.

1. Die Muster KAMPF GEGEN X, KAMPF MIT X und KAMPF DEM X können grundsätzlich denselben propositionalen Gehalt aufweisen. Trotzdem finden sich klare Präferenzen, in welchen Kontexten (mit welchen X) die Muster verwendet werden. Diese Präferenzen verändern sich diachron und unterscheiden sich synchron (z. B. bezüglich Textsorte, kommunikativer Funktion etc.). In welcher Form und syntaktischen Umgebung sie erscheinen, ist abhängig von Diskursen.
2. Für *Kampf gegen (den) Terror\** und *Kampf dem Terror\** gelten in den untersuchten Daten relativ eindeutige Präferenzen für bestimmte kommunikative Zwecke.
3. In den untersuchten Daten wird der Ausdruck *Kampf gegen (den) Terror* weit häufiger verwendet, als die ebenfalls möglichen Alternativen *Kampf dem Terror* oder *Kampf mit dem Terror*. Die Präferenz für eine Variante ist diskursiv begründet. Die vermehrte Verwendung von *Kampf dem Terror* würde z. B. auf einen veränderten Status von *Terror* schließen lassen, einen ähnlichen Status, wie er heute z. B. auch für *Krebs* oder *Stau* (*Kampf dem Krebs/Stau*) gilt. Der propositionale Gehalt GEGEN DEN TERROR KÄMPFEN bliebe zwar bestehen, die Konnotation wäre jedoch von der heutigen verschieden.

Die Analyse der Kontextualisierungsprofile der drei Muster KAMPF GEGEN X, KAMPF MIT X und KAMPF DEM X fördert also nicht nur thematisch, sondern auch pragmatisch definierte Kontexte hervor. So erscheint die Parole *Kampf dem Terror* oft in Leserbriefen und weist dort eine argumentative Funktion auf. Diese typischen pragmatischen Funktionen gelten aber oft, das haben die Analysen ebenfalls gezeigt, nur für bestimmte Textsorten und können nicht unbedingt generalisiert werden. Der Schlüssel, um die diskursiven Formierungen freizulegen, liegt in der Suche nach Veränderungen in den Kontextualisierungsprofilen.

#### 13.1.4 Was ist wert, gezählt zu werden? *Die Zahl der X*

Die Mehrworteinheit *Zahl der auf* ist typisch für den Zeitraum 2003–2005 (vgl. Zeile 67 in Tabelle 13.1 auf Seite 210). Es ist auf den ersten

Blick nicht deutlich, weshalb dies der Fall sein soll. Im Folgenden versuche ich deshalb, das zu *die/Die Zahl der X*<sup>15</sup> generalisierte Phänomen zu analysieren.

Die Wendung erscheint z. B. in folgenden Belegen:

- (36) **Die Zahl der** von Nordkorea provozierten Zwischenfälle in der demilitarisierten Zone hat neuerdings stark zugenommen. Die Politik des in Engpässe geratenen Regimes im Norden scheint auf neue Direktarrangements mit den USA abzu zielen. *Neue Zürcher Zeitung vom 18. Mai 1995, Ressort ‚Ausland‘: „Koreas Wiedervereinigung in politischer Sicht IPI-Kongress erstmals in Seoul“.*
- (37) Die EU-Kommission vertritt die Auffassung, dass die Häufung der Fälle in Frankreich eine Folge davon sei, dass die Regierung in Paris schon vorzeitig und in grösserem Ausmass Schnelltests verordnet habe. In Frankreich stieg **die Zahl der** BSE-Fälle von 31 im vergangenen Jahr auf 99 in diesem. Frankreich verdiene Lob für seinen verantwortlichen Ansatz nach dem Motto „Wer sucht, der findet“, hiess es in Brüssel. *Neue Zürcher Zeitung vom 16. November 2000, Ressort ‚Ausland‘, ap Associated Press: „Die EU-Kommission für umfangreiche BSE-Tests. Verhalten Frankreichs als Vorbild“.*

Die beiden Belege zeigen exemplarisch zwei unterschiedliche Verwendungsweisen: In Beleg 36 wird keine konkrete Zahl des Vorkommens von X (*von Nordkorea provozierte Zwischenfälle*) genannt. 56 % der 191 Belege entsprechen diesem Fall, wobei natürlich im unmittelbaren Kontext vor oder nach dem Gliedsatz, der den Ausdruck *die Zahl der* enthält, konkrete Zahlen genannt werden können. Anders in Beleg 37, wo konkrete Zahlen des Anstiegs von X (*BSE-Fälle*) angegeben sind. In 44 % der 191 Belege im Ausland-Ressort ist dies der Fall, wobei bei einem Viertel dieser 44 % die Zahlenangabe als ungefähre Angabe relativiert wird.<sup>16</sup>

<sup>15</sup> Im Folgenden wird die Groß-/Kleinschreibung nicht mehr besonders erwähnt. Es gilt aber für alle Berechnungen, dass diese ignoriert wurde.

<sup>16</sup> Dies geschieht z. B. mit Verben wie *wird geschätzt*, *schwankt*, Adjektiven wie *knapp*, *rund* oder Adverbien wie *schätzungsweise* etc.

Beiden Belegen ist gemein, dass sie eine Veränderung von X beschreiben. Es gibt aber auch Verwendungsweisen, bei denen keine Veränderung der Zahl des Gegenstandes ausgedrückt wird, wie der folgende Beleg zeigt:

- (38) Zudem sollen im Ausland geschlossene Ehen mit einem französischen Partner künftig nicht mehr automatisch anerkannt werden. Innenminister Sarkozy kündigte Einwanderungsquoten an. Es sei Sache der Nation, **die Zahl der** Migranten zu bestimmen und diese auf Grundlage ihrer Aufnahmekapazitäten auszuwählen. *Neue Zürcher Zeitung vom 30. November 2005, Ressort ‚Ausland‘, ap Associated Press: „Paris verschärft die Einwanderungsgesetze. Ankündigungen Villepins und Sarkozys“*

Im Ausland-Ressort des NZZ-Korpus entsprechen 27 % der 191 Belege diesem Typ.<sup>17</sup> Weiter gibt es vier Belege, die einen Vergleich mit *als* anstellen, wie z. B.:

- (39) Die Zahl der praktizierenden Orthodoxen liege aber bei unter zwei Prozent und sei damit noch niedriger als **die Zahl der** Mitglieder der anderen Religionsgemeinschaften, der muslimischen, evangelischen, katholischen sowie jüdischen. *Neue Zürcher Zeitung vom 30. August 2000, Ressort ‚Ausland‘, Gstrein H.: „Neue Runde im Athener Kirchenkampf. Umstrittene Gleichsetzung von Orthodoxie und Griechentum“*

Zusammenfassend zur Typologisierung der Verwendungsweisen von *die Zahl der* lässt sich vermuten, dass der Ausdruck ein Indikator für Entitäten ist, die ‚zählenswert‘ sind. Zwar wird nicht immer eine konkrete Zahl genannt, die Entitäten eignen sich aber grundsätzlich dazu, gezählt zu werden.

Wie Tabelle 13.9 auf Seite 233 zeigt, liegen die Frequenzen von *die Zahl der* im Ressort ‚Ausland‘ in der Zeit nach dem Jahr 2000 nur auf Basis der Zählung von Wörtern, nicht aber auf Basis der Zählung von Artikeln, signifikant höher als vor dem Jahr 2000. Dies gilt zwar auch für das ganze Korpus. Stärker schlägt sich jedoch eine

<sup>17</sup> Ebenfalls als Verben, die eine Veränderung ausdrücken, wurden gezählt: *begrenzen*, *eingrenzen* und *beschränken*.

Umverteilung der Mehrwortgruppe auf die Ressorts nieder. Während die Mehrwortgruppe im Ressort ‚Ausland‘ zunimmt, nimmt sie in der Sportberichterstattung ab (vgl. Tabelle 13.10 auf der nächsten Seite).

Es stellt sich nun die Frage, ob sich das Kontextualisierungsprofil von *die Zahl der* im Ressort ‚Ausland‘ in der Zeit zwischen 1995 und 2005 verändert hat. Um das Kontextualisierungsprofil zu erstellen, wurden aus dem Korpus die Füllungen für den Slot X in *die Zahl der X* extrahiert. Dabei ist X als das nächste groß geschriebene Wort nach *die Zahl der* definiert, sodass adjektivische Ergänzungen zu X miterfasst werden.

Zur Zusammenfassung dieser Ergebnisse diene eine Kategorisierung der Kontexte, wie in Tabelle 13.11 auf Seite 234 ersichtlich ist. Die Tabelle zeigt gleichzeitig die Unterschiede zwischen den zwei Zeitphasen 1995–1999 und 2001–2005.<sup>18</sup>

Es fällt auf, dass die Kategorien KRIMINALITÄT und KRIEG stark zugenommen haben, die Kategorie POLITIK hingegen abgenommen hat. Die Füllungen des Slots X in *die Zahl der X* für die drei Kategorien sind in Tabelle 13.12 auf Seite 235 ersichtlich. Die folgenden Belege geben einen Einblick in typische Kontexte dieser drei Kategorien (KRIMINALITÄT, Beleg 40; KRIEG, Beleg 41; POLITIK, Beleg 42):

- (40) Der marokkanische Justizminister hat **die Zahl der** Verhafteten vor kurzem mit 906 angegeben. Vermutlich sind es über 1000. Gegen rund 100 Islamisten laufen zurzeit Prozesse. Bisher sind 14 Todesurteile ausgesprochen worden. *Neue Zürcher Zeitung vom 20. September 2003, Ressort ‚Ausland‘, Meister U.: „Französischer ‚Emir‘ in Rabat verurteilt Terroristenprozess mit unklaren Beweisen“.*
- (41) Der stellvertretende Chef der Kaserne erklärte, die Rebellen gäben die Waffen erst nach Erhalt einer schriftlichen Erklärung des Premierministers zurück. Er müsse garantieren, dass **die Zahl der** Truppenangehörigen nicht reduziert werde. Die Gruppe aufständischer Soldaten hält sich noch immer in der

<sup>18</sup> Die Kategorisierung nur aufgrund der kontextlosen Mehrwortgruppe ist natürlich problematisch. So überschneiden sich die Kategorien teilweise und unterscheiden sich in der Spezifität (z. B. KRIEG, TOD und ISRAEL). Für die weitere Argumentation sind aber primär die Kategorien KRIEG und POLITIK entscheidend, die relativ eindeutig zu unterscheiden sind.

alle Ressorts					Ausland			
	Artikel		Wörter		Artikel		Wörter	
	#	%	#	p/Mio.	#	%	#	p/Mio.
1995	118	3,22	152	62	19	3,78	21	68
1996	116	3,13	157	64	14	3,00	14	47
1997	112	2,93	150	59	12	2,56	12	39
1998	127	3,51	161	63	21	4,47	25	87
1999	122	2,87	173	66	18	3,40	23	69
2000	127	2,82	183	67	17	3,00	24	68
2001	142	3,25	171	65	29	5,25	36	110
2002	141	3,22	188	73	21	4,05	31	109
2003	141	3,33	196	78	23	3,99	25	74
2004	153	3,68	193	80	23	4,47	25	87
2005	124	3,00	173	72	23	4,56	35	121
	$\chi^2 = 9,534$ $p > 0,10$ (nicht sign.) $V = 0,015$		$\chi^2 = 17,094$ $p < 0,10$ (signifikant) $V = 0,001$		$\chi^2 = 9,066$ $p > 0,10$ (nicht sign.) $V = 0,04$		$\chi^2 = 25,364$ $p < 0,01$ (signifikant) $V = 0,003$	

**Tabelle 13.9:** Verteilungen und  $\chi^2$ -Statistiken (Einheit: Artikel) von *die Zahl der* im ganzen Korpus und in Artikeln des Ressorts ‚Ausland‘ (für alle  $\chi^2$ -Statistiken gilt  $df = 10$ ).

Ressort	95–99 %	01–05 %	Differenz
Ausland	14,1	17,0	2,86
Magazin	7,4	8,8	1,45
Vermischtes	8,6	9,4	0,84
Lokales	13,1	13,6	0,44
Feuilleton	4,9	5,1	0,26
Wirtschaft	26,7	26,7	–0,05
Leserbriefe	1,2	0,6	–0,61
Inland	17,5	15,7	–1,79
Sport	5,0	2,9	–2,19
anderes	0,7	0,0	–0,67
unbekannt	0,8	0,3	–0,56
Total	100,0	100,0	0,00
N	595	701	

**Tabelle 13.10:** Verteilung der Artikel mit mindestens einem Vorkommen von *die Zahl der* auf die Ressorts vor und nach dem Jahr 2000.



Kontexte	1995–1999		2001–2005		+/-	Beispiel die Zahl der...
	#	%	#	%		
KRIMINALITÄT	5	5,3	21	13,8	8,6	Gefangenen, Straftaten, Todesurteile
KRIEG	6	6,3	15	9,9	3,6	Angriffe, Terroristen
STAAT	0	0,0	4	2,6	2,6	Ämter, Staatsbürgerschaften
ISRAEL	0	0,0	2	1,3	1,3	(israelischen) Siedler
NATURKATASTROPHEN	0	0,0	1	0,7	0,7	Hochwasser
TOURISMUS	0	0,0	1	0,7	0,7	Karnevalstouristen
BILDUNG	1	1,1	2	1,3	0,3	Studienplätze
WIRTSCHAFT	9	9,5	13	8,6	-0,9	Armen, Banken
MIGRATION	6	6,3	8	5,3	-1,1	Asylbewerber, Migranten
GESUNDHEIT	4	4,2	4	2,6	-1,6	Erkrankungen, Schwangerschaften
RECHT	2	2,1	0	0,0	-2,1	Prozessklagen
TOD	13	13,7	17	11,2	-2,5	Opfer, Toten
DEMOGRAPHIE	14	14,7	16	10,5	-4,2	Afghanen, Familien, Islamisten
POLITIK	13	13,7	14	9,2	-4,5	Abstimmungen, Wähler
AMBIG	14	14,7	26	17,1	2,4	Fälle, Menschen, Personen
UNVOLLSTÄNDIG	8	8,4	8	5,3	-3,2	
Total	95	100,0	152	100,0	0,0	

**Tabelle 13.11:** Kontexte von *die Zahl der X* vor und nach dem Jahr 2000 im Ressort ‚Ausland‘. Zur Kategorie AMBIG: Diese Mehrwortgruppen können sehr unterschiedlich kontextualisiert werden. Eine Bestimmung ist deshalb ohne weitere Kontextinformationen unmöglich. Zur Kategorie UNVOLLSTÄNDIG: Durch das automatische Verfahren bedingte Mehrwortgruppen ohne vollständiges Genitivobjekt zu *die Zahl der*.

Kaserne verschanzt. *Neue Zürcher Zeitung vom 21. März 2001, Ressort ‚Ausland‘, Gerdes P.: „Unruhe in der Armee von Papua-Neuguinea. Bewaffneter Widerstand gegen eine Militärreform“.*

- (42) Zur Erklärung der massiven Wahlniederlage der Unionsparteien war diesmal mit den routinemässig angewendeten Seziermethoden wenig zu gewinnen. Unbestreitbar schwindet traditionelles, voraussagbares Wählerverhalten, und **die Zahl der Wechselwähler** nimmt dauernd zu. Für professionelle Meinungsforscher, die ja mehrheitlich ein Kopf-an-Kopf-Rennen

Kategorie	1995–1999	2001–2005
KRIMINALITÄT	Anzeigen, Festnahmen*, politischen Gefangenen, nächtlich an der Polizeisperre, rechtsextremistischen Zwischenfälle	Einbrüche, Entführungsoffer, Exekutionen*, Freigelassenen, Gefangenen*, im Gefängnis*, Gewaltdelikte, (rechten) Gewalttaten, Hinrichtungen*, Polizisten*, Raubüberfälle, (registrierten) rechtsextremistischen Straftaten, Straftaten*, (vollstreckten) Todesurteile*, Verhafteten*
KRIEG	Armeebrigaden, Folterungen, zivilen Toten, UNHCR-Mitarbeiter, tschetschenischen Verteidiger, Vertriebenen*	getöteten Amerikaner, Angriffe, rund 6000 vorhandenen Atomsprengköpfe*, Attentate, Blauhelmsoldaten*, nuklearen Gefechtsköpfe, Kfor-Soldaten*, KZ-Toten, amerikanischen Militärangehörigen*, Minen*, Rebellen*, Selbstmordanschläge, Terroristen, Truppenangehörigen*
POLITIK	Abgeordneten*, Abstimmungsberechtigten, (verletzten) Demonstranten*, EU-Kommissare, Kabinettsposten, Manifestanten*, tatsächlich Stimmberechtigten*, potentiellen Wähler*, Wahlberechtigten*, Wahlkreise*, Wechselwähler	Abgeordneten*, Befürworter, ausländischen Delegierten, Demokratien, Kandidaten*, (freiheitlichen) Minister*, Nationalisten*, Nein-Stimmen*, Neonazis*, Parteimitglieder*, Protestierenden, Rechtsextremisten*

**Tabelle 13.12:** X der Wendung *die Zahl der X* vor und nach dem Jahr 2000 für die Kategorien KRIEG, KRIMINALITÄT und POLITIK im Ressort ‚Ausland‘. Mit Stern (\*) markierte Nomen kommen im Korpus zusätzlich auch in Wendungen der Art *[Zahl] ... X* vor (vgl. dazu auch Tabelle 13.13 auf Seite 237).

voraussagten, wird die Aufgabe immer schwieriger. *Neue Zürcher Zeitung vom 3. Dezember 1998, Ressort ‚Ausland‘, Kind C.: „CDU und CSU im Sog der Wahlniederlage. Suche nach Erklärungen für die Schlappe vom 27. September“.*

In Kapitel 8.2.3 habe ich bereits am Beispiel des Paradigmas STERBEN gezeigt, dass im Ressort ‚Ausland‘ seit 2001 vermehrt die Ausdrucksvarianten *(ge)tötet* und *in den Tod* verwendet werden und sich die Art und Weise, wie im internationalen, politischen Kontext über tödliche Ereignisse gesprochen wird, verändert hat. Wir sehen hier mit der Ver-

wendung von *die Zahl der X* eine ähnliche Entwicklung: Es scheint ab 2000 mehr Ereignisse im Kontext KRIEG und KRIMINALITÄT zu geben, in denen ‚zählenswerte‘ Entitäten thematisiert werden.

Doch stellt sich im Anschluss die Frage nach alternativen Ausdrucksweisen. Obwohl in der Mehrzahl der Fälle primär Veränderungen von zählbaren Entitäten thematisiert werden und oft auch gar keine konkreten Zahlen genannt werden (vgl. die Ausführungen auf den Seiten 230–231), bestünde eine alternative Ausdrucksweise darin, die Zahl der gezählten Entität einfach zu nennen:

- (43) Die von Vietnamesen abstammenden, aber schon lange in Kambodscha wohnenden Flüchtlinge erhielten von Phnom Penh die Erlaubnis zur Rückkehr. Zuletzt waren noch 4000 von ursprünglich 20 000 **Flüchtlingen** im Niemandsland, um deren Rückkehr sich die Uno bemühte. *Neue Zürcher Zeitung vom 23. Januar 1995, Ressort ‚Ausland‘, neu: „Erlaubnis für Flüchtlinge zur Rückkehr nach Kambodscha. Zwei Jahre im Niemandsland“.*
- (44) Was soll geschehen, wenn es der Schwesterpartei nicht mehr gelingen sollte, von ihrem Tiefpunkt bei **28 Prozent der bundesweiten Wählerschaft** loszukommen? *Neue Zürcher Zeitung vom 3. Dezember 1998, Ressort ‚Ausland‘, Kind C.: „CDU und CSU im Sog der Wahlniederlage. Suche nach Erklärungen für die Schlappe vom 27. September“.*
- (45) Eine internationale Geberkonferenz hatte im letzten Dezember in Tokio Aufbauhilfe von gesamthaft **522 Millionen Dollar** versprochen. *Neue Zürcher Zeitung vom 21. Oktober 2000, Ressort ‚Ausland‘, Morf U.: „Überseechinesen in Osttimor – eine Geschichte wiederholt sich. Geburtshelfer der neuen Volkswirtschaft oder reine Profitmacher?“.*

In Tabelle 13.13 auf der nächsten Seite sind die häufigsten X in Mehrwortgruppen der Art *[Zahl] ... X* aufgeführt, wobei X ein Nomen und die Zahl weder Ordinal- noch Jahreszahl ist und zwischen Zahl und X weitere Wörter, aber keine Nomen vorkommen. Die Tabelle zeigt Füllungen für X, die im Zeitabschnitt 1995–1999 oder 2001–2005 mindestens 0,3 % aller X ausmachen und zusätzlich die Differenz

[Zahl] ... X	1995–1999		2001–2005		Differenz
	#	%	#	%	
Euro	0	0,000	26	0,692	0,69
Uhr	23	0,683	41	1,091	0,41
Rebellen	3	0,089	15	0,399	0,31
Meter	10	0,297	21	0,559	0,26
Frauen	3	0,089	13	0,346	0,26
Toten	5	0,149	13	0,346	0,20
Mitglieder	6	0,178	14	0,373	0,19
Soldaten	58	1,723	70	1,863	0,14
Tote	11	0,327	15	0,399	0,07
Jahren	118	3,505	134	3,567	0,06
Personen	129	3,831	146	3,886	0,05
Demonstrant	9	0,267	12	0,319	0,05
Polizisten	9	0,267	12	0,319	0,05
Franken	19	0,564	22	0,586	0,02
Sitze	24	0,713	26	0,692	–0,02
Hektaren	17	0,505	18	0,479	–0,03
Milliarden	137	4,069	150	3,993	–0,08
Einwohner	18	0,535	17	0,452	–0,08
Einwohnern	14	0,416	11	0,293	–0,12
Stunden	35	1,040	34	0,905	–0,13
Dollar	40	1,188	37	0,985	–0,20
Seiten	16	0,475	9	0,240	–0,24
Tonnen	36	1,069	31	0,825	–0,24
Parteien	11	0,327	3	0,080	–0,25
Kandidaten	17	0,505	8	0,213	–0,29
Sitzen	19	0,564	10	0,266	–0,30
Monaten	23	0,683	14	0,373	–0,31
Flüchtlinge	20	0,594	9	0,240	–0,35
Staaten	21	0,624	10	0,266	–0,36
Jahre	72	2,138	64	1,703	–0,43
Menschen	38	1,129	22	0,586	–0,54
Prozent	450	13,365	479	12,750	–0,62
Kilometer	105	3,119	89	2,369	–0,75
Mann	68	2,020	47	1,251	–0,77
Stimmen	49	1,455	25	0,665	–0,79
Millionen	355	10,544	327	8,704	–1,84
andere	1379	40,956	1763	46,926	5,97
Total	3367	100,000	3757	100,000	

**Tabelle 13.13:** Frequenzen für X in Wendungen [Zahl] ... X im Ressort ‚Ausland‘, wobei X ein Nomen und die Zahl keine Ordinal- oder Jahreszahl ist. Zwischen Zahl und X können weitere Wörter, aber keine Nomen vorkommen. Mindestfrequenz: 0,3 % aller Wendungen in mindestens einem der Teilkorpora.

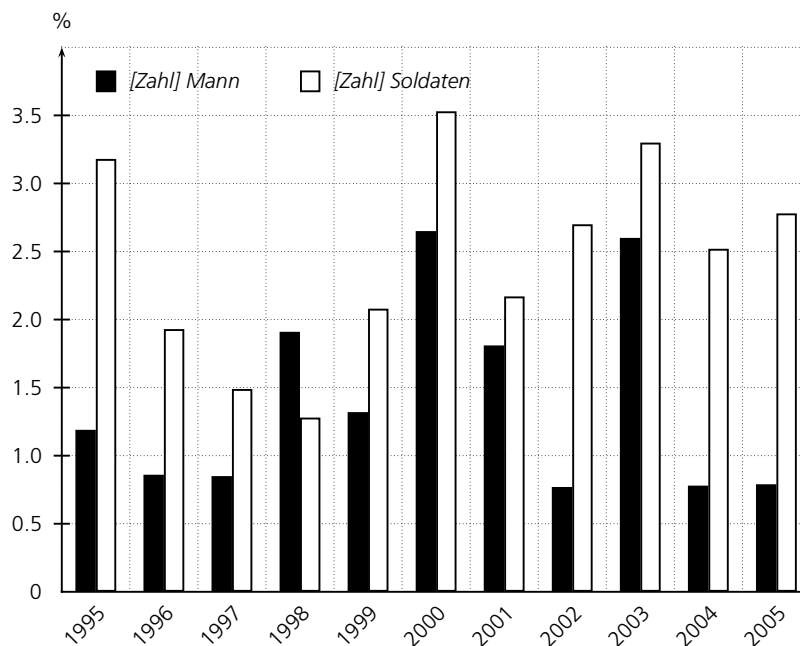
der Frequenzen in den beiden Zeitabschnitten. Nomen wie *Franken*, *Milliarden*, *Einwohner*, *Stunden* und *Dollar* kommen in beiden Zeitabschnitten häufig vor, verfügen also im Korpus über eine hohe, aber stabile Typik (vgl. Kapitel 8.2.1). Die Nomen *Euro*, *Uhr*, *Rebellen*, *Meter*, *Frauen* und *Toten* einerseits und *Millionen*, *Stimmen*, *Mann*, *Kilometer*, *Prozent* und *Menschen* andererseits sind dagegen von variabler Typik, also entweder für die Periode 2001–2005 oder 1995–1999 typisch. Neben Überraschungen (*Uhr*, *Meter*, *Frauen* bzw. *Millionen*, *Mann*, *Kilometer*, *Prozent*) ist die Typik von *Euro* für die Zeit nach 2001 durch dessen Einführung 2002 (als Buchgeld bereits 1999) erklärbar.

Interessant im Zusammenhang mit den obigen Analysen zu *die Zahl der* sind die Nomen, die dem Kontext KRIEG zugeordnet werden können und die typisch für die zweite Zeitperiode sind: *Rebellen*, *Tote(n)*, *Soldaten*. Auch bei Wendungen der Art *[Zahl] ... X* kann also eine Zunahme von kriegerischen Kontexten ab 2001 festgestellt werden.

Bedingt zur Kategorie KRIEG können *Mann* (als ‚Soldaten‘) und *Flüchtlinge* (dieser Ausdruck ist typisch für die Zeit vor 2000) zugeordnet werden. Ein Blick auf die Belege von *[Zahl] Mann* zeigt, dass in beinahe allen Fällen (insgesamt 82 Artikel im Ressort ‚Ausland‘, 183 Artikel über alle Ressorts verteilt) Soldaten gemeint sind. Beispielfhaft sei der folgende Beleg aufgeführt:

- (46) Statt wie bisher von 20 000 oder 25 000 amerikanischen Soldaten in einer Nato-Friedenstruppe ist im Verteidigungsministerium nun nur noch von **15 000 bis 20 000 Mann** die Rede. *Neue Zürcher Zeitung* vom 9. Oktober 1995, Ressort ‚Ausland‘, Schmid U.: „Fragen um den amerikanischen Bosnien-Einsatz Widerstand aus dem Kongress – Protest Buchanans“.

Die Resultate in Tabelle 13.13 auf der vorherigen Seite haben gezeigt, dass *Soldaten* typisch für die Zeit nach 2000, *Mann* typisch für die Zeit vor 2000 ist. In Abbildung 13.4 auf der nächsten Seite sind die Frequenzverläufe der beiden Ausdrücke *[Zahl] Mann* und *[Zahl] Soldaten* im Auslands-Ressort aufgeführt (vgl. für die detaillierten Daten auch Tabelle 13.14 auf Seite 240). Es wird ersichtlich, dass in den Jahren 1997–2001 und 2003 der Frequenzverlauf der beiden



**Abbildung 13.4:** Die Frequenzen der Artikel im Ressort ‚Ausland‘, die mindestens ein Vorkommen von *[Zahl] Mann* bzw. *[Zahl] Soldaten* aufweisen. Vgl. Tabelle 13.14 auf der nächsten Seite für die detaillierten Daten.

Ausdrücke etwa ähnlich ist, hingegen in den Jahren 1995, 2002 und 2005 die Frequenzen auseinanderklaffen. In diesen drei Jahren wird also die Formulierung *[Zahl] Soldaten* der Alternative *[Zahl] Mann* besonders stark vorgezogen; es wäre interessant zu sehen, ob die Präferenzen für eine der Formulierungsvarianten mit bestimmten Diskursen zusammenhängen.

Auch diese Analyse wirft eine Reihe weiterer Fragen auf, die ich an dieser Stelle nicht weiter verfolgen möchte. Bis anhin lässt sich aber Folgendes festhalten:

1. Es gibt eine Reihe von Dingen, die in der Zeitungssprache (der NZZ) zählenswert sind. Im Ressort ‚Ausland‘ sind das hauptsächlich Geld und Menschen in unterschiedlichen Zuständen und Situationen.

	<i>[Zahl] Mann</i>		<i>[Zahl] Soldaten</i>	
	#	%	#	%
1995	6	1,19	16	3,18
1996	4	0,86	9	1,93
1997	4	0,85	7	1,49
1998	9	1,91	6	1,28
1999	7	1,32	11	2,08
2000	15	2,65	20	3,53
2001	10	1,81	12	2,17
2002	4	0,77	14	2,70
2003	15	2,60	19	3,30
2004	4	0,78	13	2,52
2005	4	0,79	14	2,78
	$\chi^2 = 19,737$		$\chi^2 = 11,301$	
	$p < 0,05$		$p > 0,10$	
	(signifikant)		(nicht sign.)	

**Tabelle 13.14:** Verteilungen und  $\chi^2$ -Statistiken (Einheit: Artikel) von *[Zahl] Mann* bzw. *[Zahl] Soldaten* in Artikeln des Ressorts ‚Ausland‘ (für beide  $\chi^2$ -Statistiken gilt  $df = 10$ ). Vgl. Abbildung 13.4 auf der vorherigen Seite für eine grafische Darstellung der Verläufe.

2. In der Zeit nach 2000 bewegen sich die zählenswerten Entitäten stärker in den Kontexten KRIMINALITÄT und KRIEG, dafür weniger im Kontext POLITIK als noch vor dem Jahr 2000. Diese Verschiebungen zeigen sich sowohl im Rahmen der Formulierung *die Zahl der ... X* als auch von *[Zahl] ... X*.
3. Es scheint bei den zählenswerten Entitäten zeitgebundene Präferenzen für bestimmte sprachliche Realisierungen zu geben. So im Verhältnis der Formulierungsalternativen *[Zahl] Soldaten* vs. *[Zahl] Mann* (in der Bedeutung ‚Soldaten‘).

### 13.2 Die Inlandsberichterstattung im NZZ-Korpus

Für die Analyse der Inlandsberichterstattung dienen als Grundlage die Listen berechneter Mehrworteinheiten dieses Ressorts, die für die Perioden 1995–1997 bzw. 2003–2005 signifikant sind (vgl. Tabellen A.3 auf Seite 345 und A.4 auf Seite 347). In Tabelle 13.15 auf der nächsten Seite ist eine manuelle Auswahl dieser Listen wiedergegeben;

	G <sup>2</sup>	MWE 1995–1997		G <sup>2</sup>	MWE 2003–2005
1	962,54	Ja Ja Ja	25	–56,77	vor den Medien
2	57,43	in bezug auf	26	–41,43	der der SVP
3	38,70	im Zweiten Weltkrieg	27	–19,95	Kritik der der
4	36,20	der im Zweiten	28	–35,29	in Bezug auf
5	26,22	Schweiz im Zweiten	29	–21,48	die Ausdehnung die
6	31,21	der im Weltkrieg	30	–29,15	Bund und Kantone
7	26,22	Schweiz im Weltkrieg	31	–26,09	der Vereinten Nationen
8	32,46	Einheit der Materie	32	–16,88	der neuen EU-Staaten
9	32,46	Jugend ohne Drogen	33	–15,34	Ausdehnung der auf
10	31,21	Verhandlungen mit EU	34	–21,48	Personenfreizügigkeit auf
11	29,96	der Telecom PTT			die
12	27,46	mit gegen Stimmen	35	–19,95	Strom ohne Atom
13	24,97	wird mit Stimmen	36	–19,95	im Zuge der
14	22,47	die des neuen	37	–19,95	folgt der Mehrheit
15	18,73	ein zwischen und	38	–19,95	eine Reihe von
16	17,48	im Sinne des	39	–15,34	ebenso wie der
17	17,48	in der Praxis	40	–18,41	Stimmen der Mehrheit
18	17,48	der bilateralen Verhandlungen	41	–18,41	Referendum gegen die
19	17,48	Millionen Franken im	42	–16,88	zur Bekämpfung der
20	20,25	mit Millionen Franken	43	–16,88	die innere Sicherheit
21	16,23	Rolle der im	44	–16,88	die embryonalen Stammzellen
22	47,44	die Rolle der	45	–15,34	sich nicht nur
23	16,23	das Ergebnis der	46	–15,34	die Wirtschaft und
24	16,23	im Kampf gegen	47	–15,34	der flankierenden Massnahmen
			48	–15,34	der Eröffnung des

**Tabelle 13.15:** Typische Mehrworteinheiten im Ressort ‚Inland‘ in den Zeitperioden 1995–1997 (links) und 2003–2005 (rechts). Manuelle Auswahl für die weitere Analyse (ausführlichere Variante im Anhang A.2, vollständige Version: <http://www.bubenhofer.com/korpusanalyse/>).

die darin enthaltenen Mehrworteinheiten untersuche ich im Folgenden detaillierter.

### 13.2.1 Grobanalyse

Die in Tabelle 13.15 dargestellte Auswahl der für die frühere oder spätere Zeitperiode signifikanten Mehrworteinheiten widerspiegeln in erster Linie politische Aspekte der Schweiz. Symptomatisch dafür steht *Ja Ja Ja* (1) an erster Stelle als typischer Ausdruck des Zeitraums 1995–1997. Es handelt sich dabei um den Bestandteil eines ‚Parolenspiegels‘, der die Parolen von Parteien und Verbänden vor einer Abstimmung zusammenfasst.



Doch relevanter sind Ausdrücke, womit politische Prozesse beschrieben werden, wie *Einheit der Materie*<sup>19</sup> (8), *Verhandlungen mit EU* (10), *mit gegen Stimmen* (12), *Ausdehnung der Personenfreizügigkeit* (33f.), *Stimmen der Mehrheit* (40) oder *Referendum gegen die* (41).

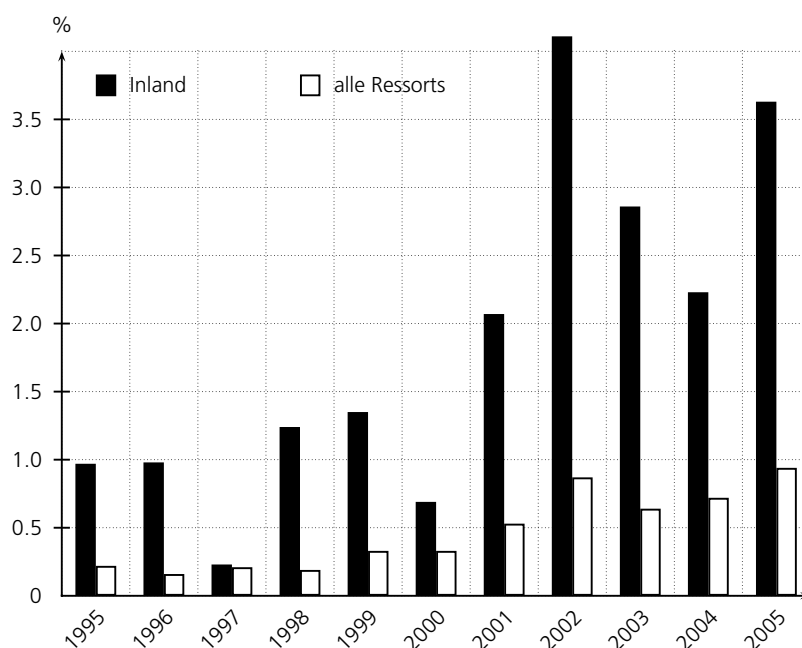
Daneben fällt der Themenkomplex die Rolle der *Schweiz im Zweiten Weltkrieg* (3ff.) auf, aber auch Ausdrücke, deren Zeittypik unklar scheint, wie *vor den Medien* (25) oder *im Zuge der* (36).

### 13.2.2 Routinisierung der medialen Selbstreflexion: *vor den Medien*

Der Ausdruck *vor den Medien* erfährt im Zeitraum von 1995–2005 sowohl im Ressort ‚Inland‘, als auch über alle Ressorts hinweg eine stetige Zunahme (vgl. Tabelle 13.16 auf Seite 244 und Abbildung 13.5 auf der nächsten Seite). Der Kontext des Ausdrucks ist typischerweise folgender Art:

- (47) Daneben gibt es auch eine informelle Zusammenarbeit, die aber, wie EJPD-Generalsekretär Armin Walpen **vor den Medien** erklärte, in jedem Einzelfall vom Departementschef bewilligt werden muss. *Neue Zürcher Zeitung vom 10. November 1995, Ressort ‚Inland‘, Lautenschütz R.: „Dreifache Aufgabe der Bundespolizei“.*
- (48) Sie [Biobäuerin Regina Fuhrer, Präsidentin der Bio Suisse (Vereinigung Schweizer Biolandbau-Organisationen), NB] persönlich wie auch der Bio-Suisse-Vorstand könnte sowohl mit einem Ja wie mit einem Nein leben. Mit oder ohne UHT-Milch – an den Grundsätzen des Biolandbaus werde sich nichts ändern, betonte sie am Dienstag **vor den Medien**. *Neue Zürcher Zeitung vom 27. März 2002, Ressort ‚Inland‘, Rosenberg M.: „Heisse Köpfe um ultraheisse Milch. Bioprodukte im Clinch mit neuen Konsumgewohnheiten“.*

<sup>19</sup> Es handelt sich hierbei um ein Kriterium, das bei Initiativen erfüllt sein muss, damit diese zur Abstimmung gelangen können.



**Abbildung 13.5:** Die Frequenzen der Artikel im Ressort ‚Inland‘ und in allen Ressorts, die mindestens ein Vorkommen von *vor den Medien* aufweisen. Vgl. Tabelle 13.16 auf der nächsten Seite für die detaillierten Daten.

Selbst wenn der Ausdruck auf *die/den Medien* oder gar *Medien* verkürzt wird, ist eine signifikante Zunahme der Verwendung sichtbar.<sup>20</sup>

Wie die Belege 47 und 48 oben zeigen, scheint die Formulierung meistens dem Kontext einer Pressekonferenz zu entstammen. Im Korpus (Ressort ‚Inland‘) gibt es zwölf Belege, in denen *Presse-/Medienkonferenz* oder *Medienorientierung* auch wörtlich genannt werden, z. B. im folgenden Artikel:

- (49) In der Schweiz ist eine dreijährige Kampagne gegen die Kinderpornographie im Internet lanciert worden. Im Visier der

<sup>20</sup> Für *die/den Medien* bewegt sich die Frequenz aller Artikel des Ressorts ‚Inland‘ 1995 von 2 % (= 10 Artikel) auf 7 % (= 34 Artikel), wobei für Korrelation zwischen Jahr und Frequenz gilt:  $\chi^2 = 33,729$ ,  $df = 10$ ,  $p < 0,001$  (signifikant). Für *Medien* ist im selben Zeitraum eine Zunahme von 4 % (= 20) auf 9 % (= 44) zu beobachten, wobei hier gilt:  $\chi^2 = 20,106$ ,  $df = 10$ ,  $p < 0,05$  (signifikant).

	<i>vor den Medien</i>			
	Inland		alle Ressorts	
	#	%	#	%
1995	4	0,97	8	0,22
1996	4	0,98	6	0,16
1997	1	0,23	8	0,21
1998	5	1,24	7	0,19
1999	6	1,35	14	0,33
2000	3	0,69	15	0,33
2001	9	2,07	23	0,53
2002	19	4,11	38	0,87
2003	14	2,86	27	0,64
2004	11	2,23	30	0,72
2005	17	3,63	39	0,94
	$\chi^2 = 37,557$		$\chi^2 = 69,165$	
	$p < 0,001$		$p < 0,001$	
	(signifikant)		(signifikant)	

**Tabelle 13.16:** Verteilungen und  $\chi^2$ -Statistiken (Einheit: Artikel) von *in den Medien* in Artikeln des Ressorts ‚Inland‘ und in allen Ressorts (für alle  $\chi^2$ -Statistiken gilt  $df = 10$ ). Vgl. Abbildung 13.5 auf der vorherigen Seite für eine grafische Darstellung der Verläufe.

Aktion stehen sowohl potenzielle Täter wie auch Jugendliche und Eltern, wie es am Donnerstag an einer **Pressekonferenz** in Bern hiess. Die Polizei erhält verschiedene Kommunikationsmittel in die Hand.

[...]

Nun müsse mit zusätzlichen Präventionsschritten die Öffentlichkeit weiter aufgeklärt werden und auf die suchtähnliche Dynamik hingewiesen werden, die bei einem übermässigen Konsum von einschlägigen Angeboten entstehen könne, hiess es **vor den Medien**. *Neue Zürcher Zeitung vom 9. September 2005, Ressort ‚Inland‘, ap Associated Press: „Kampagne gegen Kinderpornographie Internet im Zentrum“.*

Allerdings hat die Nennung von *Presse-/Medienkonferenz* oder *Medienorientierung* im Untersuchungsraum signifikant abgenommen.<sup>21</sup>

<sup>21</sup> Der Anteil der Artikel mit Nennung einer der Lexeme *Presse-/Medienkonferenz* oder *Medienorientierung* im Ressort ‚Inland‘ geht von gut 6 % (= 26 Artikel) 1995 auf knapp 3 % (= 14) 2005 zurück:  $\chi^2 = 19,259$ ,  $df = 10$ ,  $p < 0,05$  (signifikant). Noch signifikanter ist die Abnahme dieser Lexeme über alle Ressorts hinweg: von 3 % (= 110 Artikel) 1995 auf 2 % (= 86) 2005:  $\chi^2 = 40,011$ ,  $df = 10$ ,  $p < 0,001$  (signifikant).

	<i>vor den Medien</i>	
	St. Galler Tagblatt	Zürcher Tages-Anzeiger
1996	—	0,45 %
1997	0,06 %	0,60 %
1998	0,06 %	0,80 %
1999	0,13 %	0,88 %
2000	0,17 %	1,13 %
2001	0,17 %	—

**Tabelle 13.17:** Verteilungen (Einheit: Artikel) von *in den Medien* im ‚St. Galler Tagblatt‘ und im ‚Zürcher Tages-Anzeiger‘ auf Basis des DeReKo IDS (o.J.)-Korpus. Anzahl verfügbare Texte: St. Galler Tagblatt: 349 085; Tages-Anzeiger: 142 714.

Fasst man Artikel zusammen, die mindestens einen der Ausdrücke *vor den Medien* oder *Presse-/Medienkonferenz*, *Medienorientierung* enthalten, ergeben sich im Durchschnitt 125 Artikel pro Jahr, wobei keine signifikanten Unterschiede zwischen den Jahren festzumachen sind ( $\chi^2 = 9,91$ ,  $df = 10$ ,  $p > 0,10$ ). Es scheint also eine Tendenz zu sein, bei Berichterstattung auf Grundlage von Pressekonferenzen weniger oft explizit darauf hinzuweisen, dass eine *Pressekonferenz* diese Grundlage bildet und dafür das Sprachgebrauchsmuster X [SAGT] VOR DEN MEDIEN zu verwenden.

Ein Vergleich mit anderen Zeitungskorpora zeigt, dass der Ausdruck *vor den Medien* eine Eigenheit der Schweizer Presse zu sein scheint. Für den Vergleich dienten vier Teilkorpora des DeReKo IDS (o.J.)-Korpus aus unterschiedlichen Zeiträumen: ‚St. Galler Tagblatt‘ (1997–2001), ‚Zürcher Tages-Anzeiger‘ (1996–2000), ‚Mannheimer Morgen‘ (1985–2006) und ‚Die Presse‘ (1991–2000). Im Deutschen ‚Mannheimer Morgen‘ und der Österreichischen ‚Die Presse‘ finden sich kaum Treffer für *vor den Medien* (jeweils pro Jahr maximal 0,006 % bzw. 0,01 % der Artikel). In den beiden Schweizer Zeitungen ist jeweils im verfügbaren Zeitraum eine leichte Zunahme des Ausdrucks zu finden, wobei die Trefferzahl generell höher als in den nicht-schweizerischen Zeitungen ist (vgl. Tabelle 13.17). Zudem sind die Frequenzen im ‚St. Galler Tagblatt‘ etwas tiefer und im ‚Tages-Anzeiger‘ etwas höher als im NZZ-Korpus.

Wie kann die vermehrte Verwendung von *vor den Medien* in Zeitungsartikeln erklärt werden? Handelt es sich um eine zufällig entstan-

dene Formulierungsalternative, um eine zitierte Aussage als an einer Pressekonferenz geäußert zu kontextualisieren? Oder hat sich allgemeiner der Sprachgebrauch im Zusammenhang mit Medien verändert, weil sich der Diskurs über Medien verändert hat?

Einen Hinweis auf letztere Vermutung geben die Listen von in zwei Zeitabschnitten signifikanten Mehrworteinheiten, die das Lemma *Medien* enthalten (vgl. Tabelle 13.18 auf der nächsten Seite). Die Vielfalt der signifikanten ‚Medien-Mehrworteinheiten‘ ist im Vergleich zum Zeitraum 1995–1997 zurückgegangen und es gibt nur eine Mehrworteinheit, *in den Medien*, die in beiden Zeitabschnitten signifikant ist.

Nicht alle Mehrworteinheiten würde man einem wie auch immer zu definierenden ‚Medien-Diskurs‘ zurechnen. Es gibt einige Mehrworteinheiten die im Kontext von Zeitungsartikeln zu Pressekonferenzen zu vermuten sind: *am ... den/vor Medien*, *an einer Medienkonferenz/-orientierung*, *Freitag vor den Medien* etc. Daneben klingen aber eigentliche ‚Medien-Themen‘ an: Es geht um die *Medien der SRG*<sup>22</sup>, die *Printmedien* und den Platz der *Kultur in den Medien* in einem *verschärften Medienkampf*. Im Zentrum steht eine politische Debatte zur Kultur im öffentlich-rechtlichen Fernsehen mit einem Bericht des Bundesrates zur ‚Kultur in den Medien der SRG‘ (Bundesamt für Kommunikation 1997).

Diese Themen finden sich auch in anderen Ressorts der älteren Zeitperiode:

- (50) Und dass der Medienunterricht in den Gymnasien so marginal erscheint, kann durchaus in der Hoffnung begründet sein, wenn in der Schule wahre humane und wissenschaftliche Kultur vorgelebt werde, kämen die Schülerinnen und Schüler weniger in Versuchung, sich der „**Kulturlosigkeit** der modernen Medien“ auszusetzen. Im Gegensatz dazu gehen die Vertreter eines zeitgemässen Medienunterrichts davon aus, dass die Schülerinnen und Schüler auch in bezug auf Massenkommunikationsmittel „alphabetisiert“ werden sollen, dass sie lernen sollen, mit den Massenmedien bewusst und kritisch um-

<sup>22</sup> Die SRG ist die ‚Schweizerische Radio- und Fernsehgesellschaft‘, die öffentlich-rechtliche Rundfunkanstalt.

MWE (Medien) 1995–1997		#	MWE (Medien) 2003–2005		#
1	Analyse Medienereignissen der	4	32	am den Medien	7
2	dass der Printmedien	3	33	am vor Medien	6
3	den elektronischen Medien	3	34	an einer Medienkonferenz	8
4	den Medien der	5	35	an einer Medienorientierung	3
5	den Medien SRG	4	36	der den Medien	6
6	der Medien an	3	37	der der Medien	3
7	der Medien die	4	38	der Medien und	3
8	der Printmedien eine	3	39	die Medien der	3
9	die den Medien	3	40	Freitag den Medien	3
10	die in Medien	3	41	Freitag vor Medien	3
11	die Medien zu	3	42	in den Medien	12
12	die Zuger Medien	3	43	Medien den der	3
13	in den Medien	16	44	Meinungsmonopols der Medien	3
14	in elektronischen Medien	3	45	und der Mediennutzung	3
15	in Medien der	5	46	vor den Medien	37
16	in Medien SRG	4	47	welschen Medien der	3
17	Kultur den Medien	4			
18	Kultur in Medien	4			
19	Kultur Medien der	5			
20	Kultur Medien SRG	4			
21	Massnahmen gegen Mediener- zeugnisse	3			
22	Medien der SRG	4			
23	Medienereignissen der Nach- kriegszeit	4			
24	Medienereignissen der und	4			
25	Medienereignissen der Zwischen-	4			
26	Medienereignissen in der	4			
27	Medienkampf auf Anzeigerebe- ne	4			
28	Medienkampf auf die	4			
29	und den Medien	3			
30	Verschärfter Medienkampf auf	4			
31	von Medienereignissen der	4			

**Tabelle 13.18:** Signifikante Mehrworteinheiten mit dem Bestandteil *Medien* im Ressort ‚Inland‘ in den Zeitperioden 1995–1997 und 2003–2005.

zugehen. *Neue Zürcher Zeitung* vom 26. Januar 1995, Ressort ‚Schule und Erziehung‘, Jürg Scheuzger, Kantonsschule Zug: „Knöpfe in der Leitung. Die Massenkommunikationsmittel und das Gymnasium“.

In der Phase 2003–2005 scheinen diese Debatten über Kultur in den Medien abgeflaut und, obwohl das Lemma *Medien* häufiger wird (vgl. Fußnote 20 auf Seite 243), entwickelt sich ein anderer Medien-Sprachgebrauch. Dieser neue Medien-Sprachgebrauch äußert sich z. B. in der Berichterstattung in neuen Formulierungen, in denen weniger betont wird, dass eine *Pressekonferenz* stattfand, als dass Akteure *vor den Medien* sprachen. Während demzufolge in der Zeitung bereits in der ersten Phase die Thematisierung einer Selbstreflexion feststellbar ist, hat sich diese in der zweiten Phase im Sprachgebrauch der Zeitung niedergeschlagen. Es ist nicht mehr wichtig, dass die Medien thematisiert werden, sondern wie über sie gesprochen wird.

Dieser letzte Schritt der Argumentation, die ‚Routinisierung der medialen Selbstreflexion‘, ist aufgrund der vorgelegten Daten gewagt. Es müsste mit den diskursanalytischen Standardverfahren weitergearbeitet werden, um bessere Evidenz für diese These zu gewinnen.

### 13.2.3 Das Sorgenbarometer auf der Textoberfläche: *Kampf gegen X, Bekämpfung von X*

Die Mehrworteinheit *im Kampf gegen* ist für den Zeitraum 1995–1997, die Einheit *zur Bekämpfung der* für den Zeitraum 2003–2005 typisch (vgl. Zeilen 24 bzw. 42 in Tabelle 13.15 auf Seite 241). Auf die unterschiedlichen pragmatischen Funktionen von *Kampf gegen* und Varianten, sowie der zugehörigen Kontextualisierungsprofile bin ich bereits in Kapitel 13.1.3 eingegangen. Doch möchte ich hier letzteren Aspekt nochmals für das Ressort ‚Inland‘ aufnehmen und mit der Analyse des Ausdrucks *zur Bekämpfung der* ergänzen.

Die X für KAMPF GEGEN ... [AKKUSATIVOBJEKT] bzw. BEKÄMPFUNG DER/VON ... [DATIVOBJEKT] ergeben eine Liste von Bezeichnungen für Probleme, die im NZZ-Korpus thematisiert werden. Zu den häufigsten X gehören: Kriminalität, Terrorismus, Verbrechen, Arbeitslosigkeit, Geldwäscherei, Kriminalität, Korruption und Armut (vgl. Tabelle 13.19 auf der nächsten Seite). So z. B. in den folgenden Belegen:

Rang	#	MWE	KAMPF GEGEN ... X	BEKÄMPFUNG ... X
1	13	Kriminalität	✓	✓
2	10	Terrorismus	✓	✓
3	9	Verbrechen	✓	✓
4	7	Arbeitslosigkeit		✓
5	6	Geldwäscherei	✓	✓
6	5	Korruption	✓	
7	4	Armut	✓	✓
8	3	Schattenwirtschaft		✓
	3	Rassismus	✓	✓
	3	Mafia	✓	
9	2	Zollbetrug	✓	✓
	2	Windmühlen	✓	
	2	Steuerhinterziehung	✓	✓
	2	Seuche		✓
	2	Schwarzarbeit		✓
	2	Missbräuchen	✓	✓
	2	Luftverschmutzung		✓
	2	Krankheit	✓	✓
	2	Katastrophen		✓
	2	Initiative	✓	
	2	Gewalt	✓	✓
	2	Einkaufszentrum	✓	
	2	Analphabetismus	✓	✓

**Tabelle 13.19:** Die X in den Mustern KAMPF GEGEN ... X bzw. BEKÄMPFUNG ... X mit Mindestfrequenz 2 im Ressort ‚Inland‘.

- (51) Die Kommission schloss sich aber jenen Kritikern an, welche die **Bekämpfung** der organisierten Kriminalität dem Staatsschutz entziehen wollen, weil sie Kompetenzprobleme mit der Zentralstelle zur **Bekämpfung** des organisierten Verbrechens beim Bundesamt für Polizeiwesen befürchten. *Neue Zürcher Zeitung vom 15. Mai 1996, Ressort ‚Inland‘, Lautenschütz R.: „Nationalratskommission beschränkt Staatsschutz. Organisiertes Verbrechen nicht im Aufgabenbereich“.*
- (52) Der Bundesrat will den **Kampf gegen die Korruption** verstärken, jedoch keine gesetzgeberischen Schnellschüsse und internationalen Alleingänge vornehmen. *Neue Zürcher Zeitung vom 11. Dezember 1996, Ressort ‚Inland‘, ap Associated Press: „Keine Schnellschüsse im Kampf gegen Korruption“.*



Die Füllungen für X im Muster KAMPF GEGEN ... X oder BEKÄMPFUNG ... X weisen teilweise Präferenzen für das eine oder andere Muster auf. So ist im NZZ-Korpus zwar vom *Kampf gegen die Korruption*, nicht aber von der *Bekämpfung der Korruption* die Rede. Oder umgekehrt: Es gibt keinen *Kampf gegen die Initiative*, dafür die *Bekämpfung der Initiative*. Die Datenmengen sind zu gering, um aus dieser Beobachtung weitere Schlüsse zu ziehen, doch könnten mit der Analyse eines größeren Korpus interessante Erkenntnisse über diese Präferenzen gewonnen werden. So wäre z. B. zu fragen, ob es (sofern keine syntaktischen Restriktionen vorliegen) systematisch thematische Faktoren sind, die über diese Präferenzen entscheiden.

Die Liste der ‚Probleme‘ in Tabelle 13.19 auf der vorherigen Seite erinnern an Ergebnisse demoskopischer Untersuchungen, die der Bevölkerung eines Landes ‚den Puls fühlen wollen‘. In der Schweiz wird seit Ende der 90er-Jahre ein sog. ‚Sorgenbarometer‘ erstellt (Longchamp 2002, Longchamp/Golder 2003, 2004, 2005, Longchamp/Leuenberger 1999, 2000, 2001). Dieses hat den Anspruch „die Ist- und Trendanalyse zum politischen Problembewusstsein der Stimm- und Wahlberechtigten in der Schweiz“ (Longchamp/Leuenberger 1999, 9) zu leisten und basiert auf jährlichen repräsentativen Befragungen von jeweils ungefähr 1000 Personen aller drei Sprachregionen. Der Frage- teil zu den Problemen gliedert sich in zwei Teile: Eine erste Frage ist offen formuliert und fragt nach den fünf größten Problemen: „Welches sind heute Ihrer Meinung nach die fünf wichtigsten Probleme der Schweiz?“ (Longchamp/Leuenberger 2000, 8f.). Bei der zweiten Frage dienen die im vorherigen Jahr 30 häufigsten Antworten als Basis für eine gleichlautende, geschlossene Frage (Longchamp/Leuenberger 1999, 77).

Angesichts dieses Sorgenbarometers stellt sich die Frage, inwiefern sich die dort erfragten Probleme in der Medienwelt, oder konkret, der NZZ widerspiegeln. Da sowohl die Liste der zu bekämpfenden Probleme aus dem NZZ-Inland-Ressort, als auch das Sorgenbarometer schwergewichtig nationale, politische Themen berühren, ziehe ich als weitere Datenquelle die Wortprotokolle des schweizerischen Parlaments hinzu (Amtliches Bulletin o. J.). Dieses ‚Amtliche Bulletin‘ enthält die (leicht redigierten) Wortprotokolle von National- und Ständerat und der Vereinigten Bundesversammlung. Für die folgende

Rang	#	X
1	44	Kriminalität
2	43	Geldwäscherei
3	34	Terrorismus
4	22	Verbrechen
5	21	Armut
6	12	Korruption
	12	Doping
7	11	Pädophilie
8	10	Missbrauchs
9	9	Schwarzarbeit
	9	Rassismus
	9	Finanzierung
10	8	Ursachen
	8	Missbräuchen
	8	Krankheiten
	8	Bestechung
	8	Arbeitslosigkeit
11	7	Internetkriminalität
12	6	Wirtschaftskriminalität
	6	Rechtsextremismus
	6	Kulturgütertransfers
	6	Bombenanschläge
13	3	Windmühlen
	3	Warenschmuggels
	3	Teuerung
	3	Terrorismusfinanzierung
	3	Steuerhinterziehung
	3	Schafe
	3	Formen
	3	Feuerbrandes
	3	Erscheinungen
	3	Ausbeutung

**Tabelle 13.20:** Die X in den Mustern KAMPF GEGEN ... X bzw. BEKÄMPFUNG ... X mit Mindestfrequenz 3 im Amtlichen Bulletin des Schweizer Parlaments 1999–2003.

Analyse stehen mir die Protokolle aus dem Zeitraum Winter 1999 (6. Session) bis Herbst 2003 (4. Session) zur Verfügung. Es handelt sich dabei um 6934 Geschäfte unterschiedlicher Länge. Das Korpus umfasst insgesamt 12 635 844 Wörter.<sup>23</sup>

Tabelle 13.20 gibt die X der Muster KAMPF GEGEN ... [AKKUSATIV-OBJEKT] und BEKÄMPFUNG DER/VON ... [DATIV-OBJEKT] geordnet nach Auftretensfrequenz wieder. Wie bereits in der gleichartigen Berech-

<sup>23</sup> Protokolle des Nationalrats: 7 963 112 Wörter; des Ständerats: 4 610 060 Wörter; der Vereinigten Bundesversammlung: 62 672 Wörter.

nung im NZZ-Korpus wurden allfällige vor den Nomen stehende Adjektive ignoriert. Die häufigsten Nomen *Kriminalität*, *Geldwäsche*, *Terrorismus*, *Verbrechen*, *Armut* und *Korruption* nehmen im Korpus des Amtlichen Bulletins weitgehend dieselbe Reihenfolge ein wie im NZZ-Korpus, mit Ausnahme der *Arbeitslosigkeit*, die in der NZZ häufiger erscheint, sowie dem *Doping*, der *Pädophilie* und allgemein des *Missbrauchs*, die im Parlament frequenter sind als in der NZZ. Auf den unteren Rängen ergeben sich weitere Unterschiede, so tauchen im Parlament sehr konkrete X wie *Internetkriminalität*, *Kulturgütertransfer* oder die (schwarzen) *Schafe* auf, die in der NZZ höchstens vereinzelt erscheinen.

Doch wie decken sich nun diese X der beiden sprachlichen Muster mit dem, was unabhängig von syntaktischen Einbettungen als gesellschaftliches Problem charakterisiert wird? Tabelle 13.21 auf der nächsten Seite zeigt die durchschnittlichen Ränge der in der Zeit von 1999–2005 meistgenannten Probleme gemäß Sorgenbarometer.<sup>24</sup> Daneben werden, sofern genannt, die Ränge der entsprechenden X aus den Tabellen des NZZ-Korpus und des Amtlichen Bulletins aufgeführt. Dabei werden wörtliche von nur semantisch ähnlichen Nennungen unterschieden.

Von den 19 Problemen gemäß Sorgenbarometer werden nur 3 mit den selben Lexemen in den aus der NZZ und dem Amtlichen Bulletin extrahierten Listen genannt. Werden auch nicht-wörtliche Übereinstimmungen miteinbezogen, werden in der NZZ-Liste 9, in der Liste aus dem Amtlichen Bulletin 11 Probleme des Sorgenbarometers abgedeckt. Das ist doch eine erstaunlich hohe Präzision, wenn man bedenkt, dass die aus dem NZZ-Korpus und dem Amtlichen Bulletin generierten Listen aufgrund zweier eng definierten Syntagmen entstanden sind. Einige der Probleme des Sorgenbarometers können aus semantischen oder syntaktischen Gründen gar nicht als X in KAMPF GEGEN ... X oder BEKÄMPFUNG ... X erscheinen (*Bekämpfung der Altersvorsorge*, *Kampf gegen die soziale Sicherheit*), oder der Ausdruck wäre nicht usuell (*Kampf gegen Europa*, *Bekämpfung der AusländerInnen* etc.).

<sup>24</sup> Dabei berücksichtigte ich die pro Befragungsjahr 10 wichtigsten Probleme und berechnete ihre durchschnittlichen Ränge über die gesamte Periode.

Sorgenbarometer	Rang $\emptyset$	Rang NZZ	Rang AB
Gesundheit	1,71	~ 9 (Krankheit)	~ 10 (Krankheiten)
Arbeitslosigkeit	1,86	4	10
AHV/Altersvorsorge	3,00	–	–
Flüchtlinge/Asyl	4,14	–	–
Europa	4,50	–	–
Extremismus/Terrorismus	6,00	2	3
			~ 12 (Rechtsextremismus, Bombenanschläge)
			~ 13 (Terrorismusfinanzierung)
AusländerInnen	6,33	–	–
Bundesfinanzen	6,33	–	~ 9 (Finanzierung)
neue Armut	6,86	~ 7 (Armut)	~ 5 (Armut)
Steuern/Finanzen	7,50	~ 9 (Steuerhinterziehung)	~ 9 (Finanzierung)
			~ 13 (Steuerhinterziehung)
Globalisierung	8,00	–	–
soziale Sicherheit	8,00	–	–
Überfremdung	8,00	~ 8 (Rassismus)	~ 8 (Missbrauchs, Rassismus)
		~ 9 (Missbräuchen)	~ 10 (Missbräuchen)
Umwelt	8,00	~ 9 (Luftverschmutzung)	–
Wirtschaftsentwicklung	8,33	–	~ 13 (Steuerhinterziehung)
persönliche Sicherheit	8,50	~ 1 (Kriminalität)	~ 1 (Kriminalität)
			~ 4 (Verbrechen)
Löhne	9,00	–	–
Kriminalität/innere Sicherheit	9,67	1	1
		~ 3 (Verbrechen)	~ 4 (Verbrechen)
		~ 2 (Gewalt)	
Drogen	10,00	–	~ 10 (Doping)

**Tabelle 13.21:** ‚Probleme der Schweiz‘ gemäß Umfrage des ‚Sorgenbarometers‘ 1999–2005. In der Tabelle sind die Probleme nach dem durchschnittlichen Rang geordnet. Daten zusammengestellt aus Longchamp (2002), Longchamp/Golder (2003, 2004, 2005), Longchamp/Leuenberger (1999, 2000, 2001). Zusätzlich sind in den Spalten ‚Rang NZZ‘ die Ränge des entsprechenden Themas im NZZ-Korpus (Tabelle 13.19 auf Seite 249) bzw. in ‚Rang AB‘ des Amtlichen Bulletins (Tabelle 13.20 auf Seite 251) aufgeführt. Mit ~ wird eine nicht-wörtliche Übereinstimmung angezeigt.

Semantisch ambig ist das ‚Problem‘ *Missbrauch*, das in unterschiedlichen Kontexten verwendet wird. In den meisten Fällen wird die Wendung jedoch im Kontext ‚Asyl‘ benutzt, so z. B. im folgenden Votum:

- (53) Es geht hier um die glaubhafte Bekämpfung von **Missbrauch** zugunsten echter Flüchtlinge. Darum bitte ich Sie, zu dieser Motion konsequent ja zu sagen. *Nationalrat, Wintersession 1999, Fünfte Sitzung, 13. Dezember 1999, 14h30. 98.3426 Motion Fehr Hans. Staatliche Fürsorgeleistungen im Asylbereich. Votum von Fehr Hans (SVP, ZH).*

Ähnlich in einem Artikel der NZZ:

- (54) Sicherheit: Die Bevölkerung soll in Sicherheit und Ordnung leben können. Dies wird beträchtlich erleichtert durch die Integration der Eingewanderten. Aber ebenso wichtig ist die strikte **Bekämpfung von Missbräuchen** durch illegale Einwanderung oder Kriminalität. *Neue Zürcher Zeitung vom 6. Februar 2003, Ressort ‚Inland‘, Gnesa Eduard: „Wie viel Zuwanderung braucht die Schweiz? Für eine realistische Ausländerpolitik ohne Panikmache“.*

Das gemäß Sorgenbarometer Problem der ‚Überfremdung‘ findet keine direkte Entsprechung im NZZ-Korpus und in den parlamentarischen Debatten. Es gibt im NZZ-Korpus (Inland) nur 17 Belege für *Überfremdung* und keine Verwendungen von *Kampf gegen ... Überfremdung* oder *Bekämpfung der Überfremdung*. Viermal (und nur in den Belegen vor 2000) wird der Terminus in Anführungszeichen gesetzt:

- (55) In den seinerzeit von James Schwarzenbach aufgestellten Thesen – weniger ausländische Arbeitskräfte in die Schweiz holen, dafür Arbeitsstellen im Ausland schaffen – sieht Hans Steffen die Lösung für die im Zusammenhang mit der „**Überfremdung** der Schweiz“ entstandenen Probleme. *Neue Zürcher Zeitung vom 25. September 1995, Ressort ‚Eidgenössische Wahlen‘, Andreas Cleis: „Hans Steffen“.*

Im Jahr 2004 wird auf eine demoskopische Untersuchung verwiesen, die ‚Überfremdung‘ als Sorge der Schweizer Bevölkerung nennt:

- (56) Zwei Drittel der Befragten haben Angst vor einer **Überfremdung** der Schweiz. Auch die Wahrscheinlichkeit von Chemie- und Reaktorunfällen (49 Prozent, 2002: 41 Prozent) wird deutlich höher als früher eingestuft. *Neue Zürcher Zeitung vom 2. Dezember 2004, Ressort ‚Inland‘, Lezzi B.: „Gewalttätige Demonstrationen verunsichern. Publikation eines Univox-Trendberichts“.*

In den parlamentarischen Debatten erscheint *Überfremdung* ebenfalls erstaunlich selten: Nur 9 Belege finden sich in den Sitzungen zwischen 1999 und 2003, interessanterweise auch als deutscher Terminus in französischsprachigen Voten:

- (57) A mon avis – et il diverge en cela de celui du Conseil fédéral et des auteurs de la motion –, ce n’est pas en fixant des critères d’admission très restrictifs que l’on va mettre fin à ce que je nomme la forme la plus inhumaine de migration, c’est-à-dire la clandestinité. Ce n’est pas en fixant ces critères d’admission restrictifs que l’on va mettre fin à la peur de la „**Überfremdung**“ et aux sentiments xénophobes qui en découlent. *Nationalrat, Herbstsession 2000, Sechste Sitzung, 27. September 2000, 08h00. 00.3232 Motion Kommission-NR (00.016) (Minderheit Pfister Theophil). Stabilisierung des Ausländeranteils. Votum von Garbani, Valérie (SP, NE).*

Dies zeigt, dass ‚Überfremdung‘ ein Fahnenwort (Hermanns 1995, 164) der politischen Rechten ist (vgl. zu diesem „Kultur-Topos“ Wengeler 2005, 53), das allerdings ebenfalls Niederschlag im Sorgenbarometer gefunden hat.

Die Sorge vor Überfremdung schlägt sich aber durchaus im NZZ-Korpus und in den parlamentarischen Protokollen nieder: Einerseits im Problem *Rassismus*, das die politische Antwort auf die Angst vor Überfremdung sein könnte, andererseits durch Thematisierung von *Missbräuchen*, die, wie bereits oben ausgeführt, meist im Kontext ‚Asyl‘ geortet werden. Aus dieser Beobachtung könnte die These abgeleitet werden, dass der ‚Migrationsdiskurs‘ im politischen und

medialen Kontext keinen Sprachgebrauch mit dem Lexem *Überfremdung* aufweist, stattdessen ist darin vom *Kampf gegen Rassismus* und von der *Bekämpfung des Missbrauchs* als politische Reaktionen auf Überfremdungsängste in der Bevölkerung die Rede.

Es gibt auch einige Probleme, die zwar in den Listen aus dem NZZ-Korpus und/oder dem Amtlichen Bulletin erscheinen, nicht jedoch als Problem im Sorgenbarometer erwähnt werden. Dazu gehören insbesondere Aspekte der Wirtschaftskriminalität (NZZ: *Geldwäscherei, Korruption, Schattenwirtschaft*; Amtliches Bulletin: *Geldwäscherei, Korruption, Schwarzarbeit, Wirtschaftskriminalität* etc.).

Im (diskurs-)linguistischen Zusammenhang sind die oben dargestellten Befunde aus folgenden Gründen interessant:

1. Was gesellschaftlich als Problem definiert wird, zeigt sich zu großen Teilen auf der Textoberfläche als Füllungen von Slots in Sprachgebrauchsmustern wie KAMPF GEGEN ... X oder BEKÄMPFUNG ... X. Wenn die Analyse durch wenige zusätzliche Sprachgebrauchsmuster ergänzt würde, könnte ein ‚Sorgenbarometer‘ auch aufgrund von Zeitungskorpora oder Parlamentsprotokollen maschinell erstellt werden. Und wenn auch Unterschiede in so unterschiedlich generierten ‚Sorgenbarometern‘ hervortreten, sind diese interessant, um die verschiedenen Sprechweisen in Diskursen herauszuarbeiten, die sich in Parlamenten oder Medien zeigen.
2. Aus Sicht der Demoskopie stellt sich die Frage, wie stark die Antworten der Befragten diskursiv vorgeformt sind und sich überhaupt von der medialen Wirklichkeit unterscheiden.
3. Aus korpuslinguistischer Perspektive bleibt zu vermuten, dass in der Analyse von nicht-medialen Sprachdaten wie informellen Gesprächen, Web-Foren oder auch von nicht etablierten Medien interessante Einblicke in die gesellschaftlichen Diskurse und Problembewertungen gewonnen werden könnten, die sich stark vom klassischen Sorgenbarometer unterscheiden.

	G <sup>2</sup>	MWE 1995–1997		G <sup>2</sup>	MWE 2003–2005
1	12,48	nicht nur auf	27	–15,34	sich nicht nur
2	8,74	Die nicht nur	28	–9,78	nicht nur den
3	8,35	nicht nur in	29	–9,21	aber nicht nur
4	7,49	nicht nur Schweiz	30	–4,6	das nicht nur
5	7,49	nicht nur auch	31	–2,27	in nicht nur
6	4,99	nicht nur um	32	–1,49	ist nicht nur
7	4,99	nicht nur des	33	–0,03	nicht nur für
8	3,75	des nicht nur			
9	3,75	nicht nur mit			
10	3,75	Kantons nicht nur			
11	3,75	sieht nicht nur			
12	3,75	ging nicht nur			
13	3,75	nicht nur zu			
14	3,75	ihm nicht nur			
15	3,75	nicht nur als			
16	3,75	nicht nur von			
17	3,63	die nicht nur			
18	2,27	nicht nur sondern			
19	1,69	nicht nur die			
20	1,32	nicht nur im			
21	0,64	dass nicht nur			
22	0,23	nicht nur der			
23	0,09	der nicht nur			
24	0,04	sind nicht nur			
25	0,00	und nicht nur			
26	0,00	nicht nur und			

**Tabelle 13.22:** Typische Mehrworteinheiten aus dem Ressort ‚Inland‘ in den Zeitperioden 1995–1997 (links) und 2003–2005 (rechts) mit dem Ausdruck *nicht nur* als Bestandteil.

#### 13.2.4 Argumentieren und bewerten: *nicht nur ... sondern auch*

Nach den Berechnungen in Tabelle 13.15 auf Seite 241 ist die Mehrworteinheit *sich nicht nur* (45) im Ressort ‚Inland‘ signifikant für den Zeitraum 2003–2005. Allerdings ist beim Blick auf die Verteilung dieser Mehrworteinheit über die ganze Periode 1995–2005 nur eine schwache Korrelation zwischen Jahr und Frequenz auszumachen (vgl. Tabelle 13.23 auf Seite 259; 95 % Signifikanzniveau). Zwar ist in den Jahren 1998 und 2004 eine überdurchschnittliche Frequenz feststellbar, doch sind die Unterschiede zu den restlichen Jahren zu gering, um von einer hohen, variablen Typik dieser Mehrworteinheit sprechen zu können.

Neben der Mehrworteinheit *sich nicht nur* gibt es eine ganze Reihe weiterer Mehrworteinheiten mit dem Bestandteil *nicht nur*, wie Tabelle 13.22 zeigt. Die Belege für die Mehrworteinheiten zeigen, dass diese



oft zur mehrteiligen Konjunktion NICHT NUR ... SONDERN (... AUCH) erweitert werden kann:

- (58) Die Schweiz hatte **sich nun nicht nur** als nationale, sondern auch als soziale Bürgergesellschaft zu bewähren. *Neue Zürcher Zeitung* vom 1. März 1997, Ressort ‚Politische Literatur‘, Auswärtiger Autor: „Einigendes Feindbild – ambivalenter Wohlstand. Der Zusammenhalt der Schweiz vor und nach dem Krieg“.

Allerdings gibt es auch Belege, in denen das Muster nicht komplett realisiert wird:

- (59) Die amerikanische Seite weiss sehr wohl, dass die Hauptwirkung im rufschädigenden Effekt der Drohung und nicht im Boykott selber liegt. Würde dieser einmal irgendwo wirklich ausgesprochen, schnitten sich die Boykotteure **nicht nur** ins eigene Fleisch, die Betroffenen könnten mit Rechtsmitteln dagegen vorgehen, zum Beispiel beim Obersten Gerichtshof der USA, wo sie aller Wahrscheinlichkeit nach erst noch von der US-Regierung unterstützt würden. *Neue Zürcher Zeitung* vom 22. Mai 1998, Ressort ‚Inland‘, Frenkel M.: „Schatten des Zweiten Weltkriegs. Die hohe Kunst des Jonglierens“.

Wird das NZZ-Korpus (Ressort ‚Inland‘) nach Vorkommen von *nicht nur ... sondern (... auch)* durchsucht, ergeben sich durchwegs deutlich höhere Frequenzen als für *sich ... nicht nur*; die Signifikanz der Verteilung ist allerdings noch schwächer als beim alternativen Ausdruck (vgl. Tabelle 13.23 auf der nächsten Seite).

Interessant sind bei diesen Mustern also weniger die diachronen Veränderungen, sondern die Sprachgebrauchs-Charakteristik in synchroner Perspektive: Die Veränderungen ergeben sich weniger entlang der zeitlichen Achsen, sondern grundsätzlich (über die ganze Zeitperiode hinweg stabil) zwischen unterschiedlichen Texttypen oder Textfunktionen. So zeigt sich z. B. die spezifische pragmatische Funktion, die im argumentativen Kontext<sup>25</sup> erfüllt wird:

<sup>25</sup> Im Anschluss an die Argumentation in Kapitel 4.5 ist ein Textfragment dann ‚argumentativ‘, wenn es in die Struktur eines Toulmin-Schemas eingepasst werden kann.

	<i>sich ... nicht nur</i>	%	<i>nicht nur ... sondern</i>	%
1995	0	0,00	58	14,08
1996	2	0,49	43	10,51
1997	3	0,70	41	9,58
1998	7	1,73	50	12,38
1999	1	0,23	34	7,67
2000	1	0,23	51	11,72
2001	4	0,92	54	12,44
2002	1	0,22	38	8,23
2003	3	0,61	50	10,22
2004	7	1,42	49	9,92
2005	4	0,85	46	9,83
$\chi^2 = 18,488$ , df = 10			$\chi^2 = 16,636$ , df = 10	
p < 0,05 (signifikant)			p < 0,10 (signifikant)	

**Tabelle 13.23:** Verteilungen und  $\chi^2$ -Statistiken (Einheit: Artikel) von *sich ... nicht nur* und *nicht nur ... sondern (... auch)* in Artikeln des Ressorts ‚Inland‘.

- (60) Etwa 90 Prozent des Personals und 95 Prozent der Investitionen fließen nach wie vor in den Verteidigungsbereich. **Nicht nur** der für den Alleingang fehlenden Finanzen wegen, **sondern auch** wegen des liberalen Staatsverständnisses ist die Wehrkraft verhältnismässig und nicht nach traditionellen Mustern zu beanspruchen. *Neue Zürcher Zeitung vom 11. November 2004, Ressort ‚Inland‘, Auswärtiger Autor: „Wie andere Länder Armeeereformen machen. Ein Vergleich zwischen der Schweiz und Finnland“.*

Kern des Musters ist eine additive Konjunktion zweier Entitäten:

- (61) Um die Rechtsunsicherheit zu beseitigen, könnte die Sache bis ans Bundesgericht gezogen werden, was für Poschiavo eine sehr unglückliche Verzögerung bedeuten würde. Schliesslich darf nicht übersehen werden, dass der Strommarkt **nicht nur** in der Schweiz, **sondern** in ganz Europa gesättigt ist. Eine der grössten Unbekannten – zugleich aber auch der wichtigste Faktor – in diesem ganzen Szenario ist die Entwicklung des Strompreises. *Neue Zürcher Zeitung vom 19. Oktober 1995, Ressort ‚Inland‘, Auswärtiger Autor: „Kraftwerk-Heimfall am Bernina. Poschiavo probt den Alleingang“.*

- (62) Für Jacques Neirynck (cvp., Waadt) geht es bei der europäischen Integration **nicht nur** um die Interessen der Schweiz, **sondern** schlicht und einfach um Vernunft und Realitätssinn. *Neue Zürcher Zeitung vom 8. Juni 2000, Ressort ‚Eidgenössische Räte‘, Rosenberg M.: „Nachmittagssitzung. ‚Vom Bundesrat düpiert‘, ‚Totengräberei an der Demokratie‘“.*

In Beleg 62 werden im Sachverhalt der ‚europäischen Integration (der Schweiz)‘ die ‚Interessen der Schweiz‘ ins Verhältnis zu ‚Vernunft und Realitätssinn‘ gebracht. In Beleg 61 ist der ‚Strommarkt‘ der Sachverhalt, über den eine Aussage gemacht wird: Er ist ‚in der Schweiz‘ aber auch ‚in ganz Europa‘ gesättigt.

Interessant scheint, dass mit NICHT NUR ... SONDERN zwei Eigenschaften, Entitäten oder Sachverhalte erwähnt werden, die beide ‚Datum-Argumente‘ für einen ‚Claim‘ (in der Terminologie von Toulmin, vgl. Kapitel 4.5) liefern. Doch scheinen die beiden Argumente in vielen Belegen nicht gleichwertig zu sein. So kann in Beleg 62 die Implikatur abgeleitet werden, eine Integration der Schweiz in die EU sei zwar auch im Interesse des Landes, viel wichtiger, und der eigentliche Grund sei jedoch Vernunft und Realitätssinn. Diese Implikatur wird durch *schlicht und einfach* noch verstärkt: Der naheliegendste Grund ist Vernunft, nicht Interesseseolitik.<sup>26</sup>

Über mögliche Implikaturen kann nur spekuliert und ihre Existenz durch alternative Lesarten angezweifelt werden. Auffallend bei dem beschriebenen Muster ist jedoch dessen Verteilung über die Ressorts des NZZ-Korpus. Denn der Ausdruck *nicht nur ... sondern* findet sich überdurchschnittlich häufig in den Ressorts ‚Feuilleton‘ und ‚Magazin‘ und unterdurchschnittlich in den Ressorts ‚Sport‘ und ‚Vermischtes‘ (vgl. Tabelle 13.24 auf der nächsten Seite).

Beim Ressort ‚Magazin‘ handelt es sich jedoch um einen Sammeltopf unterschiedlicher Ressorts (vgl. Kapitel 11). Ein detaillierterer Blick ist deshalb nötig: Wie Tabelle 13.25 auf Seite 262 zeigt, kommt der Ausdruck häufig in feuilletonistischen Ressorts (‚Zeitfragen‘, ‚Literatur und Kunst‘, ‚Film‘, ‚Phono-Spektrum‘, ‚Zürcher Kultur‘) vor.

<sup>26</sup> Auch in der Grammatikschreibung wird die unterschiedliche Gewichtung der beiden Sachverhalte betont; nach der verneinten (oder eingeschränkten) Aussage nach *nicht nur* dient *sondern* „dem Ausdrücken, Hervorheben einer Verbesserung, Berichtigung“ dieser Aussage (Duden 1999, <sup>2</sup>sondern‘).

Ressort	<i>nicht nur ... sondern</i>	
	#	%
Feuilleton	953	17,19
Magazin	410	13,91
Inland	514	10,54
Ausland	546	9,63
Leserbriefe	94	9,20
Lokales	368	6,41
Wirtschaft	678	6,36
Sport	284	5,51
Vermischtes	116	3,82
unbekannt	13	17,11
anderes	20	16,39
$\chi^2 = 893,205$ , $df = 10$ $p < 0,001$ (signifikant)		

**Tabelle 13.24:** Verteilungen und  $\chi^2$ -Statistiken (Einheit: Artikel) von *nicht nur ... sondern* über die Ressorts des NZZ-Korpus. Absolute Frequenzen und am jeweiligen Ressort prozentuale Anteile der Artikel.

Daneben stechen Ressorts wie ‚Forschung und Technik‘ und ‚Automobil‘ hervor, die naturwissenschaftlich und technisch geprägt sind, oder ‚Fokus der Wirtschaft‘, ‚politische Literatur‘, ‚Mensch und Arbeit‘ und ‚Medien und Informatik‘ als Ressorts, in denen geistes-, sozial- oder wirtschaftswissenschaftliche Themen aufgegriffen werden. Hinter dem Ressort ‚Eidgenössische Räte‘ verbergen sich die Verhandlungsberichte der Debatten im National- und Ständerat.<sup>27</sup>

Allen diesen Ressorts gemein ist, dass sie tendenziell längere Texte enthalten. Die durchschnittliche Anzahl Wörter liegt bei den Treffern von *nicht nur ... sondern* bei 1115, während sie im ganzen Korpus bei 623 liegt. Weiter kann vermutet werden, dass in den betroffenen Ressorts eher essayistische, argumentative Texte zu finden sind, als das in den Ressorts ‚Vermischtes‘ und ‚Sport‘ der Fall ist, die nur wenige Treffer aufweisen.

Die bisher genannten Belege (58) können zwar alle als mehr oder weniger argumentativ klassifiziert werden, doch gilt das nicht für alle

<sup>27</sup> Mit der Frühlingssession 2007 (ab 6. März 2007) schafft die Neue Zürcher Zeitung ihre eigenen Verhandlungsberichte der parlamentarischen Debatten ab und verweist stattdessen auf das Amtliche Bulletin, das online abrufbar ist (Amtliches Bulletin o.J.). Im Untersuchungszeitraum 1995–2005 werden die Verhandlungsberichte jedoch lückenlos publiziert.

Originalressort	<i>nicht nur ... sondern</i>	
	#	%
Zeitfragen	28	43,75
Literatur und Kunst	135	39,02
Frankfurter Buchmesse	14	38,89
Sonderbeilage	13	37,14
Film	44	21,46
Forschung und Technik	92	21,15
Fokus der Wirtschaft	14	20,29
Mensch und Arbeit	16	20,25
Alpinismus	10	19,61
Automobil	22	19,13
Politische Literatur	22	17,46
Lebensart	18	17,14
Phono-Spektrum	15	16,67
Feuilleton	481	15,78
Eidgenössische Räte	31	14,22
Zürcher Kultur	63	14,03
Tourismus	43	12,32
Medien und Informatik	39	11,71
Fernsehen (Text)	22	10,09
unbekannt	320	12,14

**Tabelle 13.25:** Verteilungen (Einheit: Artikel) von *nicht nur ... sondern* über die Originalressorts des NZZ-Korpus. Absolute Frequenzen und am jeweiligen Ressort prozentuale Anteile der Artikel. Es werden nur Ressorts berücksichtigt, in denen die Anzahl der Artikel  $\geq 20$  und die Trefferanzahl  $\geq 10$  ist.

Belege des Ausdrucks; eine Minderheit der Belege hat eher deskriptiven Charakter, wie das folgende Beispiel zeigt:

- (63) Denn nicht die Art der Krankheit spiele für die Berechtigung zum Bezug einer IV-Rente eine Rolle, sondern nur, ob die Person ganz oder teilweise arbeitsunfähig sei. Der Bundesrat will die Revision der IV aber **nicht nur** über eine Anpassung der Ausgaben, **sondern auch** über neue Einnahmen zum Erfolg führen. Er schlägt deshalb vor, die Mehrwertsteuer um 0,8 Prozent zu erhöhen. *Neue Zürcher Zeitung vom 12. November 2005, Ressort ‚Inland‘, hof: „Gegensteuer bei der Invalidenversicherung. Arbeitgeber wollen Schleudertrauma ausklammern“.*

In argumentativen Zusammenhängen ist die nach *sondern ...* aufgeführte Einheit oft stärker gewichtet als die erste Einheit nach *nicht*

*nur* ... Welche Einheit nun die wichtigere ist, hängt vom spezifischen Kontext ab. Allerdings ist zu vermuten, dass für gewisse Themen zu gewissen Zeiten Konsens darüber herrscht, wie die Gewichtung sein muss. Daneben, dass die Füllungen in den Slots im Ausdruck *nicht nur* ... *sondern auch* ... per se in zeitlicher Abhängigkeit stehen, müsste besonders auch die Position der Füllungen zeitabhängig sein. Der folgende Beleg zeigt dies:

- (64) Als soziales Phänomen ist die organisierte Kriminalität erkannt, aber als rechtliche Kategorie bleibt sie schwer zu definieren. Eine Gefahr in internationalem Massstab geht **nicht nur** von nuklearer Proliferation und Terrorismus aus, **sondern** ebenso von der organisierten Kriminalität, wenn sie mit Waffen handelt, den Drogenhandel oder illegale Migration organisiert und in grossem Umfang Geldwäscherei betreibt. Vieles davon geschieht mit Duldung oder sogar Unterstützung durch skrupellose Regierungen, wobei eine ideologische Komponente wohl nur in Einzelfällen ausschlaggebend ist. *Neue Zürcher Zeitung vom 7. Juni 2001, Ressort ‚Ausland‘, Hauptmann W.: „Neue Dimensionen der organisierten Kriminalität“.*

In diesem Beleg werden die Gefahren von *nuklearer Proliferation und Terrorismus* der *organisierten Kriminalität* untergeordnet; eine Aussage, die nach den Anschlägen vom 11. September 2001 unwahrscheinlicher geworden ist.<sup>28</sup> Von den 7 Belegen im NZZ-Korpus für *nicht nur* ... *Terror/terror* ... *sondern* datieren zwar 3 vor und 4 nach dem 11. September 2001, 2 der 4 führen den Wortteil *Terror/terror* jedoch auch nach dem *sondern* (wörtlich und anaphorisch, vgl. Belege 65 und 66) und einer der Belege referiert auf 1952, den Zeitpunkt der im Beleg erwähnten ‚Slansky-Prozesse‘ (Beleg 67):

- (65) In der gemeinsamen Erklärung wurde die derzeit grösste Gefahr im Bereich des Terrorismus beim Namen genannt: Indiens Nachbar Pakistan. Islamabad wurde aufgefordert, **nicht nur** den „grenzüberschreitenden Terrorismus“ zu unterbinden, **sondern** überhaupt das Netzwerk des Terrors im eigenen

<sup>28</sup> Die Aussage scheint dagegen typisch für die Zeit vor dem 11. September 2001 zu sein, in der (*organisierte*) *Kriminalität* häufig thematisiert wird und auch als wichtiges zu bekämpfendes Problem angesehen wird, wie Kapitel 13.2.3 zeigte.

Land aufzulösen. *Neue Zürcher Zeitung* vom 6. Dezember 2002, Ressort ‚Ausland‘, Imhasly B.: „Indien und Russland wollen enger kooperieren. Präsident Putin beendet seinen Besuch in Delhi“.

- (66) Al-Kaida stelle immer noch eine grosse Gefahr für die Staatengemeinschaft und auch für die Schweiz dar, teilte das Eidgenössische Justiz- und Polizeidepartement mit. Unter das Verbot fallen **nicht nur** sämtliche Aktivitäten der Terrororganisation, **sondern** auch alle Aktionen, die ihrer Unterstützung dienen. *Neue Zürcher Zeitung* vom 25. November 2005, Ressort ‚Inland‘, ap Associated Press: „Verbot der Kaida um drei Jahre verlängert“.
- (67) Rückblickend findet Rybar, dass die Zeit der Slansky-Prozesse **nicht nur** vom bekannten stalinistischen Terror – einem transparenten, aber wirkungsvollen Mechanismus der Unterdrückung und der Machtsicherung – gekennzeichnet war, **sondern** auch von einer gehörigen Portion Antisemitismus. *Neue Zürcher Zeitung* vom 5. Juli 2003, Ressort ‚Zeitfragen‘, Schmid U.: „Ein Prager Überlebenskünstler. Erinnerungen des jüdischen Publizisten Ctibor Rybar“.

Während die beiden Belege 65 und 66, die nach dem 11. September 2001 entstanden sind und auch auf eine Zeit nach diesem Zeitpunkt referieren, Terrorismus stark gewichten, indem *Terror\** im zweiten (und auch im ersten) Slot von *nicht nur ... sondern auch ...* erscheint, ist dies bei einem der Belege, der im Jahr 2004 publiziert wurde, nicht der Fall:

- (68) Ausgehend vom eigentlichen Sinn der Pressefreiheit – einerseits Schutz der individuellen Freiheit, seine Meinung frei zu äussern, andererseits Verbot für den Staat, die Medien zu kontrollieren –, analysieren die Autoren den Zustand dieses Menschenrechtes weltweit. Dabei befassen sie sich **nicht nur** mit den verschiedenen Anti-Terror-Gesetzen neueren Datums, **sondern** unter anderem auch mit der Frage, ob sich die Medienfreiheit globalisieren lässt, ob sie zum Beispiel mit afri-

kanischen oder muslimischen Kulturen kompatibel sei. *Neue Zürcher Zeitung vom 3. April 2004, Ressort ‚Politische Literatur‘, Auswärtiger Autor: „Das freie Wort hat viele Feinde. Eines der gefährdetsten Menschenrechte“.*

In diesem Beleg wird also die ‚Frage nach der Globalisierung von Medienfreiheit‘ hervorgehoben und die ‚Untersuchung verschiedener Anti-Terror-Gesetze‘ etwas weniger stark gewichtet. Der Beleg widerspricht damit der These, dass nach dem 11. September 2001 *Terror\** tendenziell im zweiten Slot im Ausdruck *nicht nur ... sondern ...* verwendet wird und nicht alleine im ersten Slot erscheint. Der Beleg ist aber zusätzlich komplex, da es sich um eine Rezension eines Buches handelt, das seinerseits Gewichtungen vornimmt und die ‚Anti-Terror-Gesetze‘ zudem eine Reaktion auf Terrorismus darstellen, die ihrerseits wieder kritisiert werden können, was das rezensierte Buch wahrscheinlich macht und was der Rezensent im Beleg auch thematisiert.

Für Verwendungen von *nicht nur ... Terror ... sondern* vor dem 11. September 2001 finden sich folgende Belege (neben Beleg 64):

- (69) Zu fragen ist, weshalb von dieser Seite nicht schon früher mit mehr Nachdruck auf die Einhaltung der Menschenrechte gedrängt wurde – Anlass dazu hätte es in den letzten Jahren immer wieder gegeben, **nicht nur** wegen der notorischen Mängel im Justizwesen und der hanebüchenen Massenprozesse gegen angebliche Terrorismus-Sympathisanten, **sondern** etwa auch im Zusammenhang mit der grossflächigen Verwüstung von gegen 3000 kurdischen Dörfern in Ostanatolien. *Neue Zürcher Zeitung vom 30. Juli 1996, Ressort ‚Ausland‘, Wysling A.: „Ankaras Staatsräson unter Druck“.*
- (70) Die Spannung ist hier ganz nach innen genommen: Das Opfer [Hanns Martin Schleyer, NB] ist **nicht nur** der demütigenden Gewalt der Terroristen unterworfen, **sondern** trotz aller Hilfe der Familie hilflos den Entscheidungen der Regierung ausgeliefert, die von Anfang an nicht austauschen will. Seine Hoffnung verfällt ebenso wie die der in Stammheim einsitzenden Baader, Ensslin und Genossen, die schliesslich den einzigen Ausweg



im Selbstmord sehen, den sie dem Staat noch als Mord unterschieben wollen. *Neue Zürcher Zeitung* vom 24. Juni 1997, Ressort ‚Blick auf den Bildschirm‘, Seifert H.: „Vor 20 Jahren – Terror in Deutschland“.

Wie erwartet, wird in diesen Belegen, im Kontrast zu den Belegen nach dem 11. September 2001, *Terror\** in der ersten Position verwendet und damit die *grossflächigen Verwüstungen von gegen 3000 kurdischen Dörfern* bzw. das *hilflos den Entscheidungen der Regierung ausgeliefert* sein in den Fokus gerückt. Diese Gewichtung von *Terror\** könnte also als zeitlich begründet verstanden werden, wie dies auch im folgenden Beleg der Fall ist:

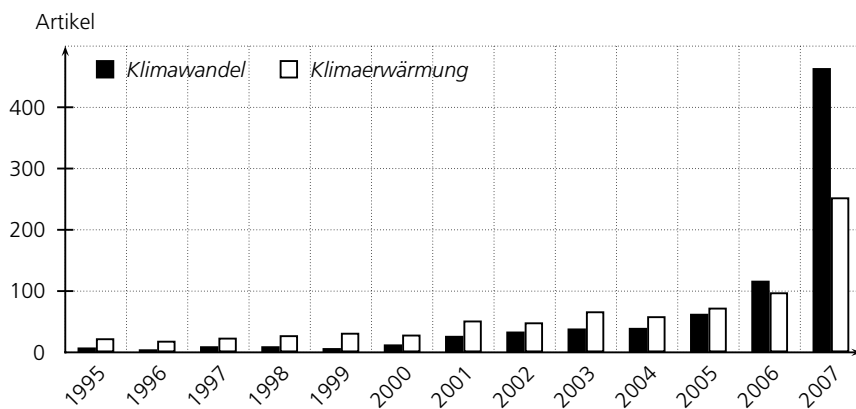
- (71) Auch gehe es dabei [beim schweizerischen CO<sup>2</sup>-Gesetz, NB] **nicht nur** um die Klimaerwärmung, **sondern auch** um Luftqualität und die Förderung neuer Technologien. Das Gesetz sei national breit abgestützt und gerade bei den letzten Energieabstimmungen in breiten Kreisen als vorbildhaft gelobt worden. *Neue Zürcher Zeitung* vom 22. November 2000, Ressort ‚Inland‘, Blattmann H.: „Plädoyer für echte Emissionsreduktionen. Philippe Roch vor der Haager Klimakonferenz“.

Die ‚Klimaerwärmung‘ wird erst in der Zeit nach 2005, dem Ende des Untersuchungszeitraums, zum breit diskutierten Thema in der Öffentlichkeit, wie Abbildung 13.6 auf der nächsten Seite zeigt.<sup>29</sup> Im Jahr 2000 war ‚Klimawandel‘ jedoch ein eher marginales Thema, so dass in Beleg 71 betont werden kann, dass der Nutzen eines CO<sup>2</sup>-Gesetzes gerade auch bei der Verbesserung der Luftqualität und der Förderung neuer Technologien liege und nicht ‚bloß‘ bei der Verhinderung der Klimaerwärmung.

Die in diesem Kapitel vorgeführten Analysen berühren sehr unterschiedliche Untersuchungsinteressen aus den Bereichen Syntax, Pragmatik/Semantik, Textlinguistik und auch Diskursanalyse. Gerade im Hinblick auf Letzteres lässt sich folgendes Fazit ziehen:

- Die Korpusanalyse hat auf ein Sprachgebrauchsmuster (NICHT NUR ... SONDERN) aufmerksam gemacht, das per se die Tendenz

<sup>29</sup> Diese Auswertung basiert nicht auf dem NZZ-Korpus, sondern auf Daten aus dem elektronischen NZZ-Archiv, das alle Ausgaben der Zeitung seit 1993 enthält.



**Abbildung 13.6:** Die Frequenzen der Artikel in der NZZ von 1995–2007 (vollständige Daten), die mindestens ein Vorkommen von *Klimawandel* bzw. *Klimaerwärmung* aufweisen.

aufweist, bestimmte pragmatische Funktionen wie Argumentation anzuzeigen. Allerdings müssen diese Funktionen nach Textsorten spezifiziert werden und sie werden auch nicht zwingend in anderen als in Zeitungstexten Gültigkeit haben.

- Die Füllungen der Slots des Sprachgebrauchsmusters sind auch aus diskursanalytischer Sicht interessant, denn sie sind diskurs- und zeitabhängig. In Kombination mit den Beobachtungen zu den pragmatischen Funktionen des Musters lassen sich Erkenntnisse über das Funktionieren eines Diskurses gewinnen. Was bedeutet es, wenn ein Thema X plötzlich stärker in argumentativen Kontexten vorkommt?
- Mit diesem einen Muster ist das Potenzial noch nicht ausgeschöpft. Gerade wenn man nach den pragmatischen Funktionen von Mustern fragt, wird es eine ganze Reihe von Sprachgebrauchsmustern geben, die im Bündel diese Funktionen wahrnehmen. Die Verwendung von *nicht nur ... sondern ...* ist ein Muster unter vielen anderen, die pragmatische Funktionen erfüllen. Die Verteilung solcher Muster im Korpus zeigt, wo welche pragmatischen Funktionen erfüllt sind und lassen Rückschlüsse auf entsprechende Textsorten zu.

G <sup>2</sup> MWE 1995–1997			G <sup>2</sup> MWE 2003–2005		
<b>Feuilleton</b>			<b>Feuilleton</b>		
1	35,60	der Kunst der	16	−36,44	zum ersten Mal
2	22,00	der Welt der	17	−29,81	es nicht mehr
3	25,27	der Jahre die	18	−23,19	der Gesellschaft und
4	21,82	in der deutschen			
<b>Wirtschaft</b>			<b>Wirtschaft</b>		
5	168,37	lag um bei	19	−71,63	Amortisation Amortisation
6	164,68	lag Uhr bei			Investition
7	78,65	lag 16 bei	20	−82,78	Die Aktien von
8	58,99	tiefen als am	21	−80,97	mit Plus von
9	100,77	um als Vortag	22	−42,04	In der Hang-Seng-Index
10	73,74	des US-Schatzamt dem	23	−46,71	In Hongkong der
11	31,95	Long Bond US-	24	−63,74	im dritten Quartal
		Schatzamt	25	−43,6	Aktuell in Franken
12	45,47	die Fusion der	26	−118,34	Yen in Franken
13	33,18	in Conf-Futures Kontrakte	27	−43,6	Marktübersicht Devisen
					Asien
			28	−35,81	der der Euro-Zone
			29	−32,7	Der notierte bei
			30	−31,14	Erwartungen der Analyti-
					ker
<b>Sport</b>			<b>Sport</b>		
14	33,33	nach wie vor	31	−64,07	ooo Zuschauer Tore
15	18,61	der Damen in	32	−26,24	von Swiss Ski
			33	−20,06	der Regular Season
			34	−18,52	der Frauen in
			35	−18,05	Los Angeles Lakers
			36	−15,43	Swiss Football League

**Tabelle 13.26:** Typische Mehrworteinheiten aus den Ressorts ‚Feuilleton‘, ‚Sport‘ und ‚Wirtschaft‘ in den Zeitperioden 1995–1997 (links) und 2003–2005 (rechts). Manuelle Auswahl für die weitere Analyse (ausführlichere Variante im Anhang A.3–A.5, vollständige Version: <http://www.bubenhofer.com/korpusanalyse/>).

### 13.3 Weitere Ressorts des NZZ-Korpus

Für weitere beispielhafte Analysen auf der Basis von Mehrworteinheiten fasse ich im Folgenden Ergebnisse aus den Ressorts ‚Feuilleton‘, ‚Sport‘ und ‚Wirtschaft‘ zusammen. Die grundlegenden Listen von berechneten Mehrworteinheiten sind in den Tabellen A.5/A.6, A.7/A.8 und A.9/A.10 auf den Seiten 350–359 wiedergegeben. Daraus extrahierte ich eine manuelle Auswahl (Tabelle 13.26).

### 13.3.1 *Zum ersten Mal und die moderne, westliche, bürgerliche Gesellschaft*

Im Ressort ‚Feuilleton‘ geht der für die Zeit 2003–2005 signifikant Ausdruck *zum ersten Mal* (16) auf eine veränderte Orthographie zurück: Zwar finden sich über den ganzen Zeitraum seit 1995 Treffer für diese Schreibung, bis 2000 aber nur in sehr geringer Zahl. Bis zu diesem Zeitpunkt dominiert die Schreibung *zum erstenmal*, die ab 2001 überhaupt nicht mehr im Korpus zu finden ist. Werden beide Schreibungen berücksichtigt, sind die Frequenzschwankungen über die Jahre im Ressort ‚Feuilleton‘ nicht signifikant.<sup>30</sup>

Der Ausdruck *der Gesellschaft und* lässt sich oft zum Muster DIE/DER [ADJEKTIV] GESELLSCHAFT verallgemeinern. Die Verteilung ist dann jedoch nicht mehr signifikant. Für die Erweiterung mit *und* gibt es zu wenig Belege im NZZ-Korpus, um quantitative Aussagen machen zu können. Hingegen fördert eine Suche nach dem allgemeineren Muster die häufigsten Adjektive hervor, die im ‚Feuilleton‘ *Gesellschaft* vorangestellt werden. Es sind dies die *moderne* (0,45 %) <sup>31</sup>, *bürgerliche* (0,38 %), *westliche* (0,23 %), *offene* (0,16 %), *demokratische* (0,16 %), *multikulturelle* (0,14 %) und *menschliche* (0,13 %) Gesellschaft.

### 13.3.2 *Es ist nicht mehr wie früher.*

Die Mehrworteinheit *es ... nicht mehr* (17) ist zwar für den Zeitraum 2003–2005 im Vergleich zu 1995–1997 signifikant, eine signifikantere Häufung zeigt sich aber im Jahr 1998.<sup>32</sup> Einen typischen Beleg für den Ausdruck bietet folgende Passage:

- (72) Wie im Spitzensport, können auch an der Universität nicht alle brillant sein. Aber es braucht einen neuen Notenmix, der dem Faktor Talent mehr Rechnung trägt. Die Schweiz kann **es sich**

<sup>30</sup> Signifikant ( $\chi^2 = 26,411$ ,  $df = 10$ ,  $p < 0,01$ ) sind die Schwankungen des Ausdrucks (unter Kombination beider Schreibungen) allerdings über alle Ressorts hinweg gesehen. Die durchschnittliche Trefferzahl (Artikel) umfasst in der Zeit vor dem Jahr 2000 2,1 %, nach dem Jahr 2000 nur noch 1,7 % der Artikel.

<sup>31</sup> In Klammern: Anteile der entsprechenden Artikel an allen Artikeln des Ressorts.

<sup>32</sup> Für die Verteilung von *es ... nicht mehr* über die ganze Untersuchungsperiode 1995–2005 gilt:  $\chi^2 = 18,345$ ,  $df = 10$ ,  $p < 0,05$  (signifikant) bei durchschnittlichen jährlichen Frequenzen von 5 Artikeln, was 1,2 % der Feuilleton-Artikel entspricht.

**nicht mehr** leisten, mögliches Potential für die neue, moderne Wirtschaftselite zu übersehen. *Neue Zürcher Zeitung* vom 3. Januar 1998, Ressort ‚Zeitfragen‘, Auswärtiger Autor: „Quer-einsteiger und Ausländer als neue Schweizer Wirtschaftselite. Globaler Austausch nationaler Führungskräfte“.

Allerdings werden mit dem Ausdruck *es ... nicht mehr* nicht nur Belege nach diesem Muster erfasst, denn *nicht mehr* ist hinsichtlich der Semantik von *mehr* ambig: Die eine Lesart begreift *mehr* als Quantifikator (Komparativ zu *viel*), der mit *nicht* negiert wird:

- (73) Hätte **es** zur Eröffnung eines extra für die Jugend konzipierten Theaterfestivals **nicht mehr** verdient als ernsthafteres Schülertheater mit gutem Sound, mässigem Tanz, einer Einführung in die Basler Jugendsprache und einer Zeigefingerwarnung vor harten Drogen? *Neue Zürcher Zeitung* vom 6. März 1998, Ressort ‚Zürcher Kultur‘, Wurzenberger G.: „Theaterfestival Blickfelder. Gestrandet auf der Kulturinsel“.

Die zweite Lesart (vgl. Beleg 72) hingegen versteht *mehr* als *nicht weiterhin*. Eine der typischsten Verwendungsweisen, die bereits als phraseologisch klassifiziert werden kann, ist wohl:<sup>33</sup>

- (74) Ich habe das Gefühl, wenn du hierbleibst, wer weiss, was passieren würde, wenn du zum Beispiel dein jüngeres Ich triffst. – Aber, willst du wirklich weitermachen? Poopay: Ach, **es ist natürlich nicht mehr wie früher**. Eigentlich bin ich längst aus der Mode. Heutzutage ist alles Technologie, CS. Ruella: CS? Poopay: Cyber Sex. Die Sendung mit der Maus für Erwachsene. *Neue Zürcher Zeitung* vom 30. Dezember 1998, Ressort ‚Zürcher Kultur‘, Muscionico D.: „Vor der Premiere. Ladylike britisch – die Damen des Hauses Ayckbourn“.

<sup>33</sup> Der zitierte Zeitungartikel ist höchst artifiziell, da er ein fingiertes Gespräch in der Zukunft zwischen zwei Figuren eines Theaterstücks wiedergibt. Dass die Autorin darin Phraseologismen als Stilmittel verwendet, ist deshalb nicht überraschend. Die Einleitung des Artikels lautet: „Wir sitzen an der Bar des Regal-Hotels in London und schreiben das Jahr 2014. Poopay, eine Spezialsex-Beraterin ohne Alter, und Ruella, eine elegante Dame Mitte 60, lassen die Korken knallen. Sie trinken auf Alan Ayckbourn, den Autor des Stücks ‚Doppeltüren‘, das an Silvester im Schauspielhaus gezeigt wird und dem sie ihre Existenz verdanken. In der Tarnung eines Klub-Sandwichs protokolliert Daniele Muscionico das Gespräch, das so – oder anders – hätte gehalten werden können“.

Verwendungen des Ausdrucks mit der Lesart *nicht weiterhin* scheinen mir sehr interessant für diskursanalytische Fragestellungen zu sein. Dies besonders dann, wenn analysiert wird, welcher Sachverhalt jeweils als *nicht mehr* eine bestimmte Eigenschaft aufweisend beschrieben wird. Es kann vermutet werden, dass mit dieser Wendung oft Topoi (vgl. Kapitel 4.5) angezeigt werden. So beispielsweise im folgenden Beleg, bei dem Kultur als den wirtschaftlichen Interessen nachgestellt verstanden wird:

- (75) Auf der höchsten Ebene dieser Mediengiganten geht **es längst nicht mehr** um Literatur, die allenfalls noch ein Nebengeschäft darstellt, sondern vor allem um die Übertragungsrechte für grosse Sportereignisse, die ihrerseits wieder hohe Werbeeinnahmen garantieren. *Neue Zürcher Zeitung vom 21. Februar 1998, Ressort ‚Literatur und Kunst‘, Auswärtiger Autor: „Partnerschaft im Literaturmarkt? Der Boom der spanischen Erzählliteratur“.*

Oder der Sachverhalt der ‚global vernetzten Welt‘ wird als Grund für eine veränderte Kunst angesehen – und ein ähnlicher Sachverhalt gilt als Begründung für ein verändertes politisches Verhalten:

- (76) Die global vernetzte Welt ist heute – scheint **es** – mit den Mitteln der Malerei **nicht mehr** zu fassen. *Neue Zürcher Zeitung vom 31. Juli 1999, Ressort ‚Feuilleton‘, Frehner M.: „Auf Schmetterlingsflügeln. Bonnard in Martigny – eine Ausstellung mit Fragezeichen“.*
- (77) In einer Demokratie und im Internetzeitalter ist **es** heute in Korea **nicht mehr** möglich, sich hinter dem einst mit Absicht errichteten Tabu zu verschanzen. Als das Parlament der Forschungsstelle keine finanzielle Unterstützung gewähren wollte, wurde sofort im Internet erfolgreich eine Sammelaktion gestartet. *Neue Zürcher Zeitung vom 14. Dezember 2005, Ressort ‚Feuilleton‘, hns: „Ein historisches Tabu wankt. Die Kollaboration mit Japan während der Kolonialzeit gerät ins Blickfeld“.*

Wenn eine musterhafte Verwendung des Sachverhalts ‚global vernetzte Welt‘ in argumentativen Kontexten nachgewiesen werden kann,

Feuilleton		Ausland		Inland	
%		%		%	
7,30	achtziger Jahre	3,37	neunziger Jahre	2,46	neunziger Jahre
7,09	sechziger Jahre	2,72	achtziger Jahre	1,56	achtziger Jahre
6,92	siebziger Jahre	1,32	siebziger Jahre	1,29	siebziger Jahre
5,08	fünfziger Jahre	1,08	sechziger Jahre	0,90	sechziger Jahre
4,82	neunziger Jahre	0,69	fünfziger Jahre	0,68	dreissiger Jahre
3,67	zwanziger Jahre	0,32	vierziger Jahre	0,64	fünfziger Jahre
3,60	dreissiger Jahre	0,30	dreissiger Jahre	0,23	zwanziger Jahre
1,89	vierziger Jahre	0,19	zwanziger Jahre	0,14	weniger Jahre
0,10	weniger Jahre	0,04	weniger Jahre	0,04	vierziger Jahre
				0,02	einiger Jahre
				0,02	achtundsechziger Jahre
N = 3919		N = 5670		N = 4878	

**Tabelle 13.27:** Prozentuale Anteile des Musters [DEKADE] JAHRE an allen Artikeln des jeweiligen Ressorts (‘Feuilleton’, ‘Ausland’ und ‘Inland’) im NZZ-Korpus 1995–2005.

z. B. angezeigt durch X IST NICHT MEHR Y, spricht das für die Kategorisierung von DIE GLOBAL VERNETZTE WELT VERÄNDERT X als Topos.

Das NZZ-Korpus ist zu klein, um die Verwendung dieses Ausdrucks systematisch derart auszuwerten, dass Aussagen über veränderte Diskurse möglich wären. Grundsätzlich scheint das Muster X IST NICHT MEHR Y aber wichtige zeitgebundene Füllungen zu erfassen.

### 13.3.3 Blick in die Vergangenheit: *die -iger Jahre*

Der Ausdruck *der ... Jahre ... die* (Zeile 3 in Tabelle 13.26 auf Seite 268) geht in vielen Fällen auf das Muster DER [DEKADE] JAHRE zurück, z. B. auf *zu Beginn der sechziger Jahre*. Über alle Jahre hinweg betrachtet sind die Frequenzunterschiede des Ausdrucks *der ... Jahre ... die* jedoch nicht signifikant. Dafür zeigen sich bei der Reduktion auf das erwähnte Muster kleine Differenzen, wenn man nach den häufigsten Dekaden fragt, die in den Zeiträumen 1995–1998 bzw. 2002–2005 erwähnt werden. Die ersten drei Positionen sind dieselben, unterscheiden sich aber in der Reihenfolge. Bis 1998 sind die häufigsten Dekaden *siebziger*, *achtziger* und *sechziger Jahre*, ab 2002 *achtziger*, *sechziger* und *siebziger Jahre*. An vierter Position erscheinen ab 2002 die *neunziger Jahre*, die in der ersten Periode viel

niedrigfrequenter sind. In beiden Perioden folgen dann die *fünfziger*, *dreissiger* und *zwanziger Jahre* – die *vierziger Jahre* werden drei- bis viermal weniger oft erwähnt als die jeweils ersten der Listen (vgl. Tabelle 13.27 auf der vorherigen Seite). Verblüffend systematisch ist die Reihenfolge der Dekaden dieses Musters im Ressort ‚Ausland‘. Über den ganzen Untersuchungszeitraum hinweg betrachtet nehmen die Frequenzen mit den *neunziger Jahren* beginnend in der Reihenfolge von der jüngeren zur älteren Dekade kontinuierlich ab bis zu den *zwanziger Jahren*. Nicht ganz so systematisch gestaltet sich die Reihenfolge im Ressort ‚Inland‘. Im ‚Feuilleton‘ werden eher Themen, die zeitlich weiter zurück liegen, abgehandelt, als das in den anderen Ressorts der Fall ist. Am häufigsten wird jeweils auf die mindestens 20 Jahre zurückliegende Dekade referiert (vor 2000: *siebziger Jahre*; ab 2000: *achtziger Jahre*). Weiter ist interessant, dass alle Dekaden oft erwähnt werden, mit Ausnahme der *vierziger Jahre*; eine Zeit, die wahrscheinlich eher mit *die Zeit des Zweiten Weltkriegs*, *nach dem Krieg* und ähnlichen Formulierungen erfasst wird.

In den Ressorts ‚Ausland‘ und ‚Inland‘ wird offensichtlich am häufigsten auf die unmittelbare Vergangenheit referiert. Je weiter zurück eine Zeit oder ein Ereignis liegt, desto seltener wird darauf mit [DEKADE] JAHRE verwiesen.

### 13.3.4 Veränderungen in der Wirtschaftswelt

Im Wirtschaftsressort stechen neue geografische Bezeichnungen ins Auge. Statt des *US-Schatzamtes* (Zeilen 10, 11 in Tabelle 13.26 auf Seite 268) bis 1997 ist in der zweiten Zeitperiode vom *Hang-Seng-Index* in *Hongkong* und von *Devisen Asien* die Rede (22, 23, 27). Allerdings ist die Frequenz für *Hang-Seng-Index* nur 1995 unterdurchschnittlich tief; danach liegt die jährliche Frequenz jeweils bei durchschnittlich 17,6 Artikeln (1,79 % der Artikel im Ressort). Aus diesem Befund könnte die Hypothese abgeleitet werden, dass sich der Fokus der NZZ vermehrt auf das Wirtschaftsgeschehen in Asien verschoben hat. Eine Prüfung der Frequenzen des Lemmas *Asien* zeigt, dass dieses in 5,26 % der Artikel erscheint, wobei das Jahr 1998 hervorsteht, in dem 13,1 % der Artikel *Asien* enthalten. Diese Häufung ist auf die ‚Asien-Krise‘ zurückzuführen; in 36 % aller *Asien*-Belege findet sich *Asien-Krise*. Der Fokus in der NZZ liegt demnach primär auf dieser Krise und



das Vermehrte Vorkommen von Lemmata und Wortgruppen mit dem Bestandteil *Asien-* geht darauf zurück.

Auch tauchen in der neueren Zeitperiode die Währungen *Yen* und *Euro* auf (26, 28). Der erste Beleg für *Euro* taucht im Korpus im Januar 1996 auf – in Anführungsstrichen:

- (78) Wettbewerbsverzerrende Abwertungen seien nicht erwünscht. Die draussen bleibenden Währungen dürften vielmehr nur entsprechend dem Inflationsunterschied von ihrem Leitkurs zur Einheitswährung „**Euro**“ abweichen. *Neue Zürcher Zeitung vom 27. Januar 1996, Ressort ‚Wirtschaft‘, Münster P.: „Zugang zur WWU durch ein ‚immer engeres Tor‘. Die Stabilitätskriterien aus der Sicht des BIZ-Präsidenten“.*

In den folgenden Belegen wird *Euro* nicht mehr in Anführungsstrichen gesetzt und die jährliche Frequenz bewegt sich von 1,8 % über 9,3 % (2003) auf 7,1 % der Artikel.

Der *Yen* wird nur in den Jahren 1995 und 1998 überdurchschnittlich oft erwähnt (7,7 % und 8,9 % der Artikel) und kommt ansonsten in etwa 5 % aller Artikel des Ressorts vor.

Ebenfalls verschwinden in diesem Ressort die Wendungen *lag um (16 Uhr) bei, tiefer als am* und *um als Vortag* (5ff.). So kommt die Wendung *lag ... um ... bei* zwischen 1995 und 1998 in 1,1 % bis 2,5 % aller Wirtschaftsartikel vor, danach gibt es nur noch vereinzelte Treffer. Bis 1998 lässt sich die Wendung in den meisten Fällen auf die Formel ergänzen: *In Zürich lag der Dollar um 16 Uhr bei ... was gegenüber dem Vortagesschluss/Freitagsschluss von ... einer Einbusse/Verteuerung um ... gleichkommt.* Auf der anderen Seite gibt es in der Börsenberichterstattung erst ab Mai 1998 Belege für die Wendung *mit einem Plus von*. Die Berichterstattung über die Börsen ist in der NZZ also grundlegend verändert worden.<sup>34</sup>

### 13.3.5 Geografische Referenzen

In der Untersuchung zum Ausdruck *die bosnischen Serben* (Kapitel 13.1.2) habe ich bereits die Bedeutung der Verwendung von Länder-

<sup>34</sup> Auch die Börse in Zürich veränderte sich im Untersuchungszeitraum. 1995 schlossen sich die Börsen Genf, Basel und Zürich zur ‚Swiss Exchange‘ zusammen, 1996 wird der Ringhandel eingestellt. Vgl. <http://www.swiss-exchange.ch/> (14. 2. 2008).

oder Ethnienbezeichnungen dargestellt. Auch einige Mehrworteinheiten aus den Ressorts ‚Feuilleton‘, ‚Wirtschaft‘ und ‚Sport‘ beinhalten geografische und ethnische Referenzen: *in der deutschen* (4; vgl. auch Kapitel 13.3.6), *US-Schatzamt* (10f.), *in Hongkong der* (23) und *Los Angeles Lakers* (35). Daraus stellt sich generell die Frage nach geografischen Bezeichnungen im NZZ-Korpus.

Die Abbildungen 13.7 bis 13.9 auf den Seiten 276–277 zeigen anhand einer Karte, welche Staaten, Hauptstädte und Einwohner/innen in wievielen Artikeln erwähnt werden. Als Basis für die Zählungen dienten Listen aller Staaten (inkl. Adjektiven, z. B. *amerikanisch*, und Einwohner/innen-Bezeichnungen, z. B. *Amerikaner/Amerikanerin* sowie deren Flexionsformen).<sup>35</sup> Ebenfalls berücksichtigt sind die Namen aller Hauptstädte und der Kontinente.<sup>36</sup> Die Abbildungen 13.7, 13.8 und 13.9 auf den Seiten 276–277 zeigen die Frequenzen der Artikel in den Ressorts ‚Ausland‘, ‚Wirtschaft‘ und ‚Feuilleton‘, wobei die Punktgröße die zur höchsten Ortsfrequenz relative Frequenz symbolisiert.<sup>37</sup>

Nicht weiter überraschend wird im Auslandsressort am verbreitetsten über die ganze Welt berichtet. Trotzdem ergeben sich klare Schwerpunkte: Europa, Russland, der Nahe Osten und Nordamerika werden am häufigsten genannt. Im Ressort ‚Wirtschaft‘ sind es vor allem die Schwerpunkte Europa, USA und etwas weniger häufig Asien (in Asien besonders häufig Japan und China). Die Feuilletonberichterstattung ist sehr auf Europa bezogen, doch wird die USA auch oft erwähnt.

Eine andere Perspektive auf die geografischen Referenzen bieten die Abbildungen 13.10 und 13.11 auf Seite 278. In diesen Darstellungen, die ebenfalls die in Artikeln der Ressorts ‚Ausland‘ und ‚Wirtschaft‘ genannten geografischen Bezeichnungen zeigen, symbolisiert die Punktgröße den Variationskoeffizienten. Dieses Maß, das die Standardabweichung  $s$  in Relation zum Mittelwert setzt ( $s/\bar{x}$ ), ermöglicht

<sup>35</sup> Die Bezeichnung *Schweden* ist ambig: Sie wird sowohl für das Land als auch für die Bezeichnung der Bewohner Schwedens verwendet. Nach meinen Recherchen trifft das jedoch nur für diesen Fall zu und kann deshalb vernachlässigt werden.

<sup>36</sup> Vgl. für detailliertere Angaben auch die Ausführungen auf Seite 195.

<sup>37</sup> Die Karte ist in ihrer elektronischen Version interaktiv: Die Parameter ‚Jahr‘ und ‚Ressort‘ können dabei verändert werden und ein Klick auf die referenzierten Orte bietet Informationen über die genauen Frequenzen, sowie weitere statistische Angaben. Die Karte kann unter <http://www.bubenhofer.com/korpusanalyse/> eingesehen werden.



**Abbildung 13.7:** Ressort ‚Ausland‘: Nennung von Staaten, Hauptstädten und Einwohner/innen im NZZ-Korpus ( $N = 27083$ ,  $\bar{x} \geq 2$ ).



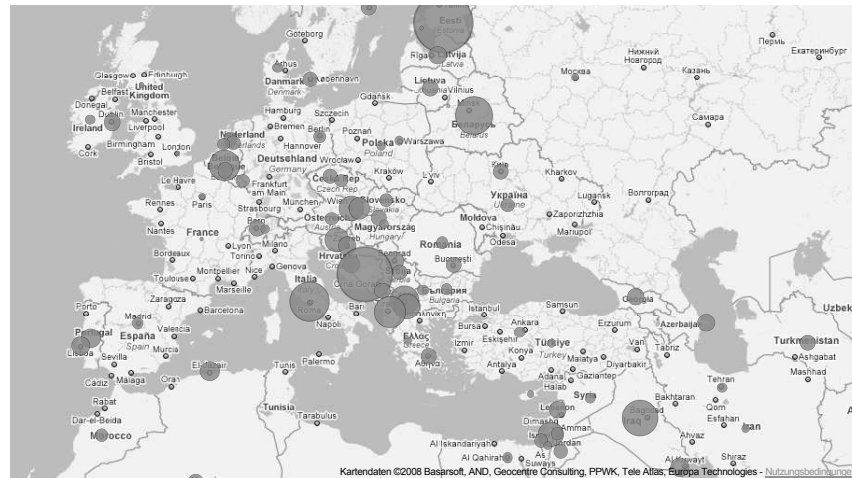
**Abbildung 13.8:** Ressort ‚Wirtschaft‘: Nennung von Staaten, Hauptstädten und Einwohner/innen im NZZ-Korpus ( $N = 30873$ ,  $\bar{x} \geq 2$ ).



**Abbildung 13.9:** Ressort ‚Feuilleton‘: Nennung von Staaten, Hauptstädten und Einwohner/innen im NZZ-Korpus ( $N = 18\,870$ ,  $\bar{x} \geq 2$ ).

einen Vergleich der Variation von verschiedenen Datensätzen, deren Frequenzen sich auf unterschiedlichen Niveaus bewegen (vgl. Kapitel 7.1). Somit zeigt die Punktgröße an, wie stark sich die Frequenzen im Zeitraum von 1995–2005 verändern.

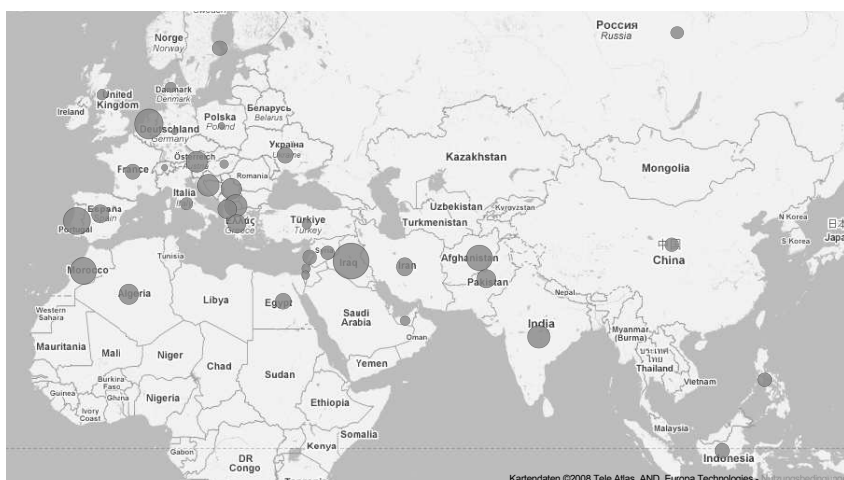
Der besseren Übersicht wegen zeigen diese Karten (Abbildung 13.10 auf der nächsten Seite und 13.11 auf der nächsten Seite) einen vergrößerten Ausschnitt Europas und des Nahen Ostens. Denn in diesem Raum ergaben sich die höchsten Schwankungen in der Erwähnung geografischer Bezeichnungen. Die Berichterstattung über die Schweiz, Deutschland, Frankreich, Spanien, Großbritannien, Polen und andere Länder mehr ist sehr konstant (detailliertere Kartenausschnitte der Abbildungen 13.7 bis 13.9 auf den Seiten 276–277 würden aber zeigen, dass die Frequenzen dieser Länder nicht alle auf gleich hohem Niveau liegen). Den größten Schwankungen ist der Balkan, aber auch Italien, Estland und der Irak unterworfen, sowie etwas weniger stark Israel (mit Gazastreifen und den palästinensischen Gebieten). Vor allem für den Balkan gelten diese starken Schwankungen jedoch nur für die Artikel im Auslandsressort (Abbildung 13.10 auf der nächsten Seite), im Ressort ‚Wirtschaft‘ (Abbildung 13.11



**Abbildung 13.10:** Ressort ‚Ausland‘: Die größten Veränderungen der Nennungen von Staaten, Hauptstädten und Einwohner/innen im NZZ-Korpus (Variationskoeffizient  $s/\bar{x}$ ,  $N = 27083$ ,  $\bar{x} \geq 2$ ).



**Abbildung 13.11:** Ressort ‚Wirtschaft‘: Die größten Veränderungen der Nennungen von Staaten, Hauptstädten und Einwohner/innen im NZZ-Korpus (Variationskoeffizient  $s/\bar{x}$ ,  $N = 30873$ ,  $\bar{x} \geq 2$ ).



**Abbildung 13.12:** Ressort ‚Ausland‘: Die größten Veränderungen der Nennungen von Einwohner/innen (Nationalitäten) im NZZ-Korpus (Variationskoeffizient  $s/\bar{x}$ ,  $N = 3205$ ,  $\bar{x} \geq 1$ ).

auf der vorherigen Seite) ist die Berichterstattung über diese Region konstanter.

Einen differenzierteren Blick auf die Welt bietet Abbildung 13.12, bei der als Basis für die Kartendarstellung nur die Frequenzen der Artikel mit Nennungen von Nationalitäten, also der Einwohnerinnen und Einwohner der einzelnen Staaten verzeichnet sind. Auch diese Karte zeigt die größten Frequenzschwankungen an und sie bietet weitere Hinweise für die Hypothese, die bereits aufgrund der Mehrworteinheit *die bosnischen Serben* in Kapitel 13.1.2 aufgetaucht ist: Ist die vermehrte Nennung von Ethnienbezeichnungen ein Indikator für kriegerische, oder zumindest problematische Kontexte?

Der Blick auf die Karte zeigt tatsächlich eine Reihe von Brennpunkten: Irak, der Balkan, Israel und die palästinensischen Gebiete, Afghanistan und Pakistan, Indien, Marokko, Algerien, Ägypten, aber auch Holland, Portugal und Spanien.<sup>38</sup> Auf der Karte nicht sichtbar sind weitere Nationalitäten: Argentinien, Peru, Kolumbien und Kanada. Die Kartendarstellung gibt also tatsächlich weitere Hinweise

<sup>38</sup> Hier sind der Lesbarkeit willen Staaten bzw. Regionen genannt; gezählt wurden aber, wie erwähnt, Einwohner/innen-Bezeichnungen, also *Iraker(n)*, *Bosnier(n)* etc.

darauf, dass vermehrte Ethnienbezeichnungen auf kriegerische Kontexte schließen lassen.

Die verwendete Methode, mit der nur nach Einwohner/innen-Bezeichnungen von Staaten gesucht wurde, ist zu vereinfachend. Damit werden nur offizielle Bezeichnungen gefunden, die sich an (offiziellen) Bezeichnungen von Staaten orientieren. Gerade in kriegerischen Konflikten, wie z. B. auf dem Balkan, sind es viele weitere Ethnienbezeichnungen, die eine Rolle spielen. So ist auf der Karte in Abbildung 13.12 auf der vorherigen Seite auch Afrika weitgehend unberührt, weil an den kriegerischen Auseinandersetzungen dort weniger ganze Nationen als vielmehr Ethnien beteiligt sind.

### 13.3.6 Der nördliche Nachbar: *Die deutsche(n) X*

Der Ausdruck *in der deutschen* (Zeile 4 in Tabelle 13.26 auf Seite 268) kommt in der Periode 1995–1997 im Ressort ‚Feuilleton‘ signifikant häufiger vor als in der Periode 2003–2005. Zieht man jedoch die Frequenzen aller elf Jahre hinzu, ist die Verteilung nicht mehr signifikant.<sup>39</sup> Zudem bewegen sich die Frequenzen auf relativ niedrigem Niveau. Allerdings ist *in der deutschen* nicht die einzige Mehrworteinheit mit dem Lemma *deutsch*, das signifikant für die Zeit von 1995–1997 ist. Tabelle 13.28 auf der nächsten Seite listet die Mehrworteinheiten auf, die sowohl für den Zeitraum 1995–1997 als auch für 2003–2005 signifikant sind und das Lemma *deutsch* enthalten.

Die Mehrheit der Ausdrücke ist eine Konstruktion der Art [DEF. ART.] DEUTSCHE(N). Fasst man die Mehrworteinheiten deshalb zu diesem Muster zusammen, ist die Korrelation zwischen Jahr und Frequenz nicht mehr signifikant.<sup>40</sup> Doch wie bereits am Beispiel von NICHT NUR ... SONDERN ... (Kapitel 13.2.4) demonstriert, sind auch bei diesem Muster die spezifischen Füllungen des Slots X in [DEF. ART.] DEUTSCHE(N) X von Interesse. Tabelle 13.29 auf Seite 282 gibt einen Überblick über die für drei Perioden typischen X.

Als ressortspezifische und nicht zeitabhängige Füllungen können die X bezeichnet werden, die in allen drei Zeitperioden vorkommen

<sup>39</sup> Im Durchschnitt enthalten 3,09 % aller Artikel den Ausdruck und es gilt:  $\chi^2 = 13,683$ ,  $df = 10$ ,  $p > 0,10$  (nicht signifikant).

<sup>40</sup> Im Durchschnitt enthalten 9,67 % aller Artikel den Ausdruck und es gilt:  $\chi^2 = 6,145$ ,  $df = 10$ ,  $p > 0,10$  (nicht signifikant).

	G <sup>2</sup>	MWE 1995–1997		G <sup>2</sup>	MWE 2003–2005
1	21,82	in der deutschen	13	–9,94	in deutscher Übersetzung
2	8,04	der die deutschen	14	–6,62	in der deutschsprachigen
3	6,89	die die deutschen	15	–6,62	für die deutschen
4	6,89	die deutsche die	16	–6,62	der deutschen die
5	4,59	mit der deutschen	17	–6,62	bei der deutschen
6	4,59	Der deutsche Film	18	–4,97	im deutschen Sprachraum
7	4,59	auf deutsch und	19	–4,97	Eine deutsche chinesisch
8	3,45	der grossen deutschen	20	–4,97	der deutschen Österreichs
9	3,45	die deutschen Truppen	21	–4,97	Jahrbuch für finnisch-
10	3,45	unter deutschen und			deutsche
11	3,45	von Können deutsche	22	–4,97	der deutschen Staaten
12	3,45	die deutschen Übergriffe	23	–4,97	der deutschen Hafiz-
					Rezeption
			24	–4,97	deutsche spricht chinesisch

**Tabelle 13.28:** Typische Mehrworteinheiten aus dem Ressort ‚Feuilleton‘ in den Zeitperioden 1995–1997 (links) und 2003–2005 (rechts) mit dem Lemma *deutsch* als Bestandteil. Auswahl der je 12 signifikantesten Mehrworteinheiten.

(ab Zeile 1): *die deutsche Literatur, Übersetzung, Sprache, Ausgabe, Geschichte, Botschaft* und *die deutschen Juden*. Allerdings sind bereits hier teilweise markante Frequenzunterschiede sichtbar: Die *deutsche Literatur* wird ab 1999 immer seltener erwähnt, dafür nimmt die Nennung von *deutsche Übersetzung* und *deutsche Ausgabe* stetig zu.

Zu diesem Bild passen Nennungen, die nur in einer der drei Zeitperioden vorkommen (ab Zeile 16). Dazu gehört *die deutsche Gegenwartsliteratur, der deutsche Buchmarkt* und *die deutsche Nachkriegsliteratur*. Die Auswertung erweckt den Eindruck einer starken Thematisierung der deutschen Literatur in der Zeit von 1995 bis 1998, die in den darauf folgenden Jahren kontinuierlich abnimmt. Stattdessen wird übersetzte Literatur erwähnt. Ein Blick in die Zahlen des Deutschen Buchhandels zeigt, dass in der Belletristik der Anteil der Übersetzungen ins Deutsche im Untersuchungszeitraum kontinuierlich von 44 % (1994) über 36,5 % (1999) auf 13,8 % (2005) zurückgegangen ist.<sup>41</sup> Die meisten Titel werden aus dem Englischen übersetzt;

<sup>41</sup> Vgl. zu diesen Zahlen Börsenverein des Deutschen Buchhandels (1995, 2000, 2006). Im Detail:

1994 umfasste die Produktion von Erst- und Neuauflagen 10 092 Titel (nur Erstauflagen: 6680), davon waren 4441 Übersetzungen, wiederum davon 3291 aus dem Englischen.

1999 umfasste die Produktion von Erstauflagen 7566 Titel, davon waren 2760 Übersetzungen, wiederum davon 1938 aus dem Englischen.



X	1995–1998		1999–2001		2002–2005	
	#	%	#	%	#	%
<i>Vorkommen in allen drei Zeitabschnitten</i>						
1 Literatur	20	9,95	6	3,73	3	1,75
2 Übersetzung	7	3,48	8	4,97	10	5,85
3 Sprache	7	3,48	6	3,73	7	4,09
4 Ausgabe	3	1,49	5	3,11	6	3,51
5 Geschichte	2	1,00	7	4,35	2	1,17
6 Juden	3	1,49	2	1,24	2	1,17
7 Botschaft	2	1,00	2	1,24	2	1,17
<i>Vorkommen in zwei von drei Zeitabschnitten</i>						
8 Schweiz	4	1,99	17	10,56		
9 Gesellschaft	2	1,00	2	1,24		
10 Künstler	2	1,00	2	1,24		
11 Kultur			6	3,73	5	2,92
12 Hauptstadt			4	2,48	2	1,17
13 Kunst			2	1,24	4	2,34
14 Presse	2	1,00			2	1,17
15 Sprachraum	2	1,00			2	1,17
<i>Vorkommen in nur einem Zeitabschnitt (Auswahl)</i>						
16 Gegenwartsliteratur	5	2,49				
17 Botschafter	4	1,99				
18 Leser	4	1,99				
19 Romantik	4	1,99				
20 Buchmarkt	3	1,49				
21 Nachkriegsliteratur	3	1,49				
22 Okkupation	3	1,49				
23 Sozialismus	3	1,49				
24 Bundeskanzler			3	1,86		
25 Intellektuellen			3	1,86		
26 Lyrik			3	1,86		
27 Regisseurin					3	1,75
28 Leserschaft					2	1,17

**Tabelle 13.29:** Vorkommen (Einheit: Artikel) von X des Musters [DEF. ART.] DEUTSCHE(N) X im Ressort ‚Feuilleton‘ im Vergleich zwischen den Zeitperioden 1995–1998, 1999–2001 und 2002–2005. Absolute Frequenzen und am Muster prozentuale Anteile der X. Es werden nur X berücksichtigt, bei denen die Anzahl  $\geq 1\%$  der Artikel mit dem Muster ist. Lesehilfe: Der Ausdruck *die deutsche Literatur* kommt im Zeitraum 1995–1998 20 Mal vor. Das bedeutet, dass in 9,95 % aller Vorkommen des Musters [DEF. ART.] DEUTSCHE(N) X in dieser Zeitperiode X den Inhalt *Literatur* hat.

doch auch hier ist eine Abnahme von 32 % (1994) über 25 % (1999) auf 7,7 % (2005) gemessen an der ganzen Belletristik-Titelproduktion zu beobachten. Wenn also im Feuilleton vermehrt Übersetzungen thematisiert werden, geht das nicht mit einer entsprechenden Mehrproduktion an Übersetzungen im Buchhandel einher – im Gegenteil. Die Zahlen würden erwarten lassen, dass in neuerer Zeit viel häufiger von deutschsprachiger Literatur die Rede wäre.

Eine interessante Einzelbeobachtung ist, dass *der/die deutsche/n Leser* von 1995–1998 durch *die deutsche Leserschaft* in den Jahren 2002–2005 abgelöst wird (Zeilen 18 und 28). Dies signalisiert gegenüber einer geschlechtsneutralen Schreibung gemachte Konzessionen, wobei sich die NZZ als relativ resistent erweist, wie Tabelle 13.30 auf der nächsten Seite zeigt. Denn die Form *Leser* (inkl. Flexions- und Pluralformen) ist über die ganze Zeitperiode hinweg gesehen die verbreitetste Form. Die geschlechtsneutrale Formulierung *Leserschaft* folgt mit etwa achtmal niedrigerer Frequenz. Alle weiteren Formen sind im Korpus marginal. Darüber hinaus gibt es über den Beobachtungszeitraum keine signifikanten Frequenzveränderungen.<sup>42</sup>

Zurück zu den Füllungen des Musters [DEF. ART.] DEUTSCHEN X. Die Füllungen des Slots X ab Zeile 8 in Tabelle 13.29 auf der vorherigen Seite kommen jeweils in zwei der drei Zeitabschnitten vor. Besonders auffallend ist der Ausdruck *die deutsche Schweiz*, der besonders im Zeitraum 1999–2001 oft genannt wird, der allerdings primär auf einen sprachwissenschaftlichen Artikel zu Schweizerdeutsch und Hochdeutsch zurückgeht.<sup>43</sup> Die weiteren Ausdrücke weisen, zumindest auf der Basis Artikel, keine signifikante Korrelation mit den Zeitab-

---

2005 umfasste die Produktion von Erstauflagen 11 187 Titel, davon waren 1542 Übersetzungen, wiederum davon 861 aus dem Englischen.

Ab 1997 wurde die Zählweise von Übersetzungen geändert, indem nur noch Erstauflagen erfasst werden. Deshalb sind die Zahlen von 1994 nur bedingt mit den späteren Jahrgängen vergleichbar. Unter Einbezug der Neuauflagen bei der gesamten Titelproduktion ergibt sich jedoch eine prozentuale Vergleichbarkeit. Alle Zahlen gelten für Deutschland.

<sup>42</sup> Die NZZ publizierte einen Artikel zum Thema: Neue Zürcher Zeitung vom 25. August 2001, Ressort ‚Zeitfragen‘, Elmiger Daniel: „Liebe(r) LeserIn, chère lectrice et chère lecteur. Die sprachliche Gleichbehandlung in den Landessprachen“.

<sup>43</sup> Neue Zürcher Zeitung vom 31. Juli 1999, Ressort ‚Zeitfragen‘, Sonderegger Stefan: „Bewegung im Sprachlaboratorium. Schweizerdeutsch und Hochdeutsch aus sprachwissenschaftlicher Sicht“.

	#	⊙ #	⊙ %	$\chi^2$
Leser/n/s	667	68,27	1,68	8,681
Leserschaft	85	8	0,19	6,334
Leserinnen und Leser/n	31	2,81	0,07	8,051
Leserin	19	1,72	0,04	15,720
Leserkreis	8	0,72	0,02	5,702
Leserinnen	3	0,27	0,01	8,161
Leser(innen)	2	0,18	0,00	9,411
Total	815	82	2,02	

**Tabelle 13.30:** Vorkommen (Einheit: Artikel) verschiedener Varianten von LESER/IN im Singular und Plural im NZZ-Korpus über die Zeitperiode 1995–2005. Die Prozentwerte drücken die Anteile der Treffer an allen Artikeln aus. Der  $\chi^2$ -Wert drückt die Stärke der Korrelation zwischen Jahr und Frequenz aus; je höher  $\chi^2$ , desto stärker die Veränderung der Frequenz über die Jahre. (Für alle  $\chi^2$ -Werte gilt:  $df = 10$ ,  $p > 0,10$ , Korrelation nicht signifikant.)

schnitten auf und können demzufolge als ressortspezifisch und nicht zeitabhängig betrachtet werden.

Weiter oben bin ich bereits auf einige der Ausdrücke, die nur in einer der drei Zeitperioden vorkommen (ab Zeile 16) eingegangen. Neben der erwähnten, sind drei Nennungen von *des deutschen Sozialismus* (23) in der Zeit vor 1999 auffallend. Die Nennungen gehen auf drei Artikel zurück (Buch- und Theaterbesprechung, Nachruf).<sup>44</sup> Der Ausdruck *der/dem deutsche Bundeskanzler* (24) in den drei Artikeln von 1999 und 2000 denotiert den damaligen Kanzler Gerhard Schröder. Der Ausdruck *deutscher Bundeskanzler* scheint jedoch ansonsten im Feuilleton kaum verwendet zu werden.<sup>45</sup>

<sup>44</sup> Neue Zürcher Zeitung vom 5. Dezember 1995, Ressort ‚Feuilleton‘, Auswärtiger Autor: „Meister der querschlagenden Sätze. Ernst-Wilhelm Händlers ‚Stadt mit Häusern‘“; Neue Zürcher Zeitung vom 21. Juni 1996, Ressort ‚Feuilleton‘, Villiger B.: „Requiem für die Untoten Heiner Müllers ‚Germania 3‘: diesmal in Berlin“; Neue Zürcher Zeitung vom 7. August 1997, Ressort ‚Feuilleton‘, Lütkehaus Ludger: „Linientreuer Dissident. Zum Tod von Jürgen Kuczynski“.

<sup>45</sup> In diesem Zusammenhang interessiert die Verwendung des Ausdrucks *Gerhard Schröder* als Alternative zu *der/dem deutsche/n Bundeskanzler*: Ab 1999, seiner Amtseinsetzung als Bundeskanzler Deutschlands, wird *Gerhard Schröder* durchschnittlich in 0,2 % der Feuilleton-Artikel erwähnt; in der verkürzten Form *Schröder* in 0,6 % der Artikel. Diese Alternative wird demnach im Feuilleton häufiger verwendet als die Funktionsbezeichnung. Die Vergleichswerte für das Ressort ‚Ausland‘ liegen für die Vollform *Gerhard Schröder* bei 0,2 % und für den Nachnamen *Schröder* bei 3,8 % der Artikel (für jeweils die Amtszeit 1999–2005).

Bei den Füllungen, die nur in einer der drei Zeitperioden vorkommen, gleichzeitig aber niedrige Frequenzen aufweisen, ist Vorsicht geboten: Sie können zwar durchaus als ressorttypisch bezeichnet werden; ihre zeitliche Abhängigkeit ist jedoch oft zu wenig signifikant.

Die Untersuchung des Musters [DEF. ART.] DEUTSCHE(N) X offenbart ein statistisches Problem: Interessiert man sich nicht nur für das abstrahierte Muster, sondern für Teilmengen des Musters, also z. B. die Verteilung der Füllungen für X über die Zeit oder die Ressorts, nimmt die Trefferzahl stark ab und es ist kaum mehr möglich, statistisch abgesicherte Aussagen zu machen. Dafür müsste das Korpus weit größer sein. Trotzdem kann man von einem großen Potenzial solch ähnlich gelagerter Untersuchungen ausgehen: Die Komponenten ‚Muster‘ und ‚Slots‘ bieten die Kombination von Sprechweisen und Sprechinhalten. Äußerungen werden in ihrem (musterhaften) Kontext analysiert und nicht als kontextlose Lemmata.

### 13.3.7 Vom Schweizer Skiverband zu *Swiss Ski*

Bei den Ausdrücken in der Sportberichterstattung sind die beiden für die Zeit nach 2003 typischen Wortgruppen *von Swiss Ski* (Zeile 32 in Tabelle 13.26 auf Seite 268) und *Swiss Football League* (36) Resultat von Umbenennungen. Im NZZ-Korpus wird *Swiss Ski* erstmals im Jahr 2000 erwähnt (vorher: *schweizerischer Skiverband*), *Swiss Football League* 2003 (vorher: *Nationalliga*):

- (79) Deshalb haben die mittlere Kammer sowie die 15 Vereine der Gruppe 1 für entgangene Einnahmen sowie zusätzliche Ausgaben hohe Schadenersatzforderungen angedroht. Ähnliches behält sich auch die Nationalliga (neu **Swiss Football League**) vor, zumal die Geschichte möglicherweise noch lange nicht zu Ende ist. *Neue Zürcher Zeitung* vom 7. August 2003, Ressort ‚Sport‘, Wesbonk R.: „Renitenter FC Sion. Keine Teilnahme am Erstligabetrieb – hohe Kosten für alle Parteien“.

Die neuen englischsprachigen Namen für Sportvereine sind denn auch im Jahr 2000 Thema einer Glosse in der NZZ:

- (80) Jedes Kind in den USA weiss, dass Cheese, Chocolate und AB-BA Schweizer Erfindungen sind, frei nach dem Motto Swiss or

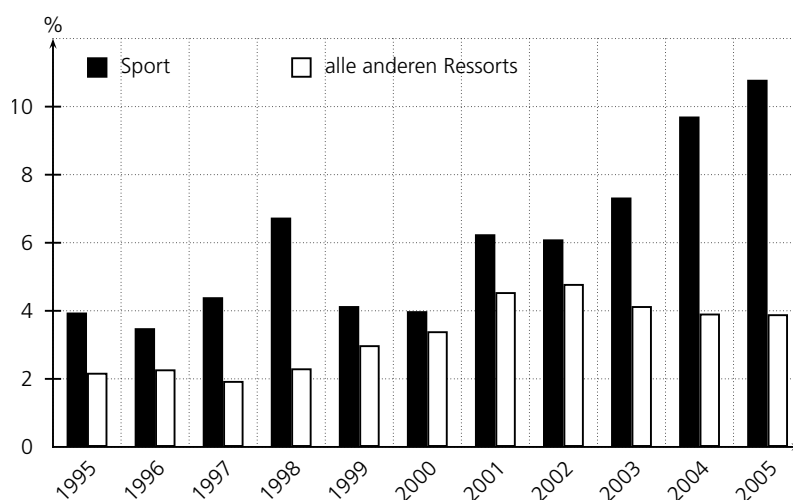
Swedish, Mädderhorn or Cape Horn, Gore or Bush – wo ist da der Unterschied! History und Schnee von gestern. Zukünftige Generationen (von Sportfans) haben's einfacher, Früh-Englisch und Early Geographie gebühre Dank: Swiss Tennis, Swiss Ski, Swiss Cycling werden zweifelsfrei identifiziert als a) Sportarten, die b) was mit Käse, Schokolade oder „Money, money, money“ zu tun haben. *Neue Zürcher Zeitung vom 22. Dezember 2000, Ressort ‚Sport‘, Fisch C.: „Marginalie. Swiss Something“.*

Die in dieser Glosse behauptete Zunahme von Bezeichnungen mit dem Bestandteil *Swiss* lässt sich statistisch tatsächlich nachweisen, wie Abbildung 13.13 auf der nächsten Seite zeigt. Allerdings nimmt die Frequenz im Ressort ‚Sport‘ erst ab 2001 deutlich zu; 1998 gab es einen ersten Ausschlag. Zum Zeitpunkt der Publikation der Glosse gibt es aber vergleichsweise wenig Vorkommen von *Swiss* in der Sportberichterstattung. Zu den häufigsten Wortkombinationen mit *Swiss* gehören (Frequenzen in Klammern): Swiss Ski (65), Swiss Olympic (64), Swiss Open (55), Swiss Indoors (46), Swiss Tennis (37) und Swiss Handball (24).<sup>46</sup>

Zum Vergleich sind in Abbildung 13.13 auf der nächsten Seite auch die Frequenzen von *Swiss* in den anderen Ressorts aufgeführt. Auch dort ist eine Steigerung von 1998 bis zum Höhepunkt 2002 zu beobachten – anschließend gehen die Frequenzen leicht zurück. Die hohen Frequenzen 2001 und 2002 sind zu einem beträchtlichen Teil auf Vorkommen von *Swissair* zurückzuführen; die Fluggesellschaft glitt im Herbst 2001 in den Konkurs und dominierte die Politik auch im darauf folgenden Jahr.<sup>47</sup>

<sup>46</sup> Zu den weiteren Bezeichnungen gehören: Swiss Team (22), Swiss-Ski (19), Swiss Cup (18), Swiss Football (17), Swiss Cycling (17), Swiss Golf (12), Swiss League (11), Swiss Derby (11), Swissauto (9), Swiss Master (8), Swiss-Meeting (7), Swiss Rookies (7), Swiss-Cup (6), Swiss Timing (6), Swiss Stars (6), Swiss Meeting (6), Swiss Badminton (6), Swiss Alpine (6), Swiss Swimming (5), Swiss Meetings (5), Swiss Unihockey (4), Swiss Inline (4), Swiss House (4), Swiss Bike (4).

<sup>47</sup> Im Jahr 2001 gibt es 96 Artikel (= 2,5 % der Artikel in diesem Jahr) mit dem Vorkommen *Swissair*, während im Zeitraum 1995–2000 die durchschnittliche Frequenz bei 29,16 Artikeln (= 0,85 %) lag.



**Abbildung 13.13:** Die Frequenzen der Artikel in den Ressorts ‚Sport‘ bzw. in allen anderen Ressorts im NZZ-Korpus, die mindestens ein Vorkommen des Wortteils *Swiss* aufweisen. Für das Ressort ‚Sport‘ gilt:  $\chi^2 = 49,898$ ,  $df = 10$ ,  $p < 0,001$  (signifikant); für alle anderen Ressorts gilt:  $\chi^2 = 114,401$ ,  $df = 10$ ,  $p < 0,001$  (signifikant). N (Sport): 5153 Artikel; N (andere Ressorts): 39690 Artikel.

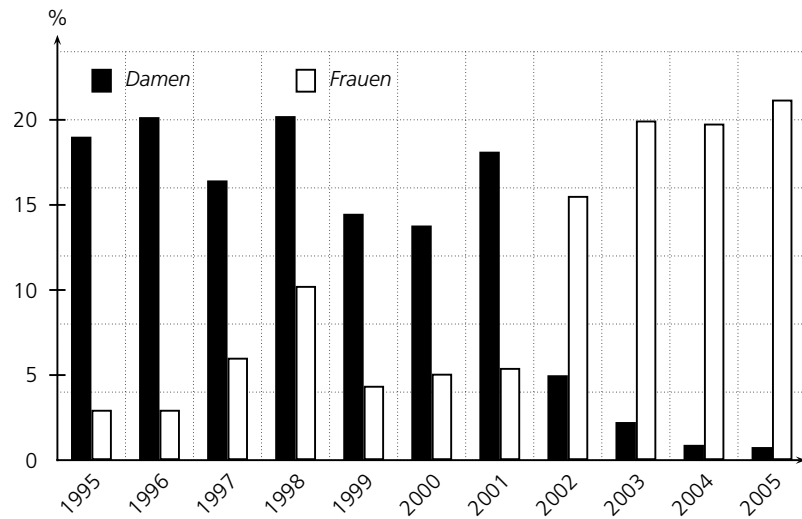
### 13.3.8 Die Damen und die Frauen im Sport

Wenn 1995 über weibliche Sportler geschrieben wird, klingt das so:

- (81) Am Tennis-Masters **der Damen** in New York ist ein deutsch-deutscher Final nach wie vor möglich. Die Weltranglisten-Erste Steffi Graf besiegte in den Viertelfinals Mary Joe Fernandez (USA/8) 6:3, 6:4, und Anke Huber setzte sich gegen die Japanerin Kimiko Date mit 3:6, 6:2, 6:1 durch. *Neue Zürcher Zeitung* vom 18. November 1995, Ressort ‚Sport‘, *Sportinformation*: „Steffi Graf und Anke Huber in New York in den Halbfinals“.

Zehn Jahre später wird das in der NZZ anders formuliert:

- (82) Die Fussballauswahl **der Frauen** verliert **in** Zug gegen Russinnen 0:2. *Neue Zürcher Zeitung* vom 2. September 2005, Ressort ‚Sport‘, cag: „Unpräzise Schweizerinnen. Die Fussballauswahl der Frauen verliert in Zug gegen Russinnen 0:2“.



**Abbildung 13.14:** Die Frequenzen der Artikel im Ressort ‚Sport‘ im NZZ-Korpus, die mindestens ein Vorkommen von *Damen* bzw. *Frauen* aufweisen. Vgl. Tabelle 13.31 auf der nächsten Seite für die detaillierten Daten.

Der Ausdruck *der Damen in* ist typisch für die Zeit von 1995–1997, während 2003–2005 der Ausdruck *der Frauen in* verwendet wird (Zeilen 15 und 34 in Tabelle 13.26 auf Seite 268). Ganz allgemein ist die Verabschiedung des Lemmas *Damen* aus der Sportberichterstattung zu beobachten: Nach dem Jahr 2001 kommt das Lemma kaum mehr vor, dafür übernimmt das Lemma *Frauen* den Platz (vgl. Abbildung 13.14 und Tabelle 13.31 auf der nächsten Seite).

Diese Umgewichtung betrifft jedoch nur die Verwendung von *Damen* bzw. *Frauen* im Ressort ‚Sport‘. Alle weiteren Ressorts zusammengekommen zeigen keine derartige Veränderung (vgl. Tabelle 13.31 auf der nächsten Seite). Bei *Damen* bewegen sich die Frequenzen relativ stabil auf eher niedrigem Niveau über die ganze Periode hinweg. Die Verwendung von *Frauen* weist sehr viel höhere Frequenzen auf, unterliegt jedoch von 1995 bis 2005 signifikanten Schwankungen und weist 1995 eine Spitze auf.

Die Verwendung von *Damen* hat im Beobachtungszeitraum, und insbesondere bis und mit 2001 in der NZZ, klar fachsprachlichen

	Sport				Andere Ressorts			
	<i>Damen</i>		<i>Frauen</i>		<i>Damen</i>		<i>Frauen</i>	
	#	%	#	%	#	%	#	%
1995	77	19,01	12	2,96	13	0,40	232	7,11
1996	75	20,16	11	2,96	27	0,81	169	5,06
1997	71	16,44	26	6,02	24	0,71	170	5,01
1998	75	20,22	38	10,24	23	0,71	213	6,56
1999	63	14,48	19	4,37	27	0,71	229	6,01
2000	76	13,79	28	5,08	22	0,56	212	5,37
2001	87	18,13	26	5,42	16	0,41	203	5,21
2002	27	4,99	84	15,53	22	0,57	162	4,21
2003	11	2,24	98	19,96	25	0,67	179	4,78
2004	5	0,90	110	19,78	15	0,42	178	4,95
2005	4	0,77	110	21,19	16	0,44	174	4,82
<b>Total</b>	571	11,08	562	10,91	230	0,58	2121	5,34
	$\chi^2 = 308,174$ $p < 0,001$ (signifikant)		$\chi^2 = 269,48$ $p < 0,001$ (signifikant)		$\chi^2 = 13,168$ $p > 0,10$ (nicht sign.)		$\chi^2 = 49,631$ $p < 0,001$ (signifikant)	

**Tabelle 13.31:** Vorkommen (Einheit: Artikel) von *Damen* bzw. *Frauen* im Ressort ‚Sport‘ und in allen anderen Ressorts des NZZ-Korpus. Die Prozentwerte drücken die Anteile der Treffer an allen Artikeln des Jahres und der Ressortauswahl aus. Vgl. Abbildung 13.14 auf der vorherigen Seite für eine grafische Darstellung.

Charakter. Eine redaktionelle Direktive scheint dafür verantwortlich zu sein, dass ab 2002 der Ausdruck *Damen* vermieden und stattdessen *Frauen* verwendet wird. Genau die gleiche Entwicklung ist im Ressort ‚Sport‘ auch mit den Lemmata *Herren* und *Männer* zu beobachten.<sup>48</sup>

Dennoch gibt es selbst 2005 im Sportressort noch vier Artikel mit Vorkommen von *Damen*. Darunter fallen zwei Artikel im Kontext ‚Schach‘; einer davon lautet:

- (83) Hier behindert der Springer die natürliche Entwicklung des **Damenläufers** nach b7, und der folgende Tausch ist für Weiss günstig. Logischer ist 13. – b6. 14. Dc2 S: c4 15. L: c4! d: c4 16. S: c4 *Neue Zürcher Zeitung vom 16. September 2005, Ressort ‚Schach‘, jen: „Kasan. Die russische Meisterschaft in alter Stärke“.*

<sup>48</sup> Die Sportverbände scheinen die Selbstbezeichnungen nur teilweise zu ändern. So führt der Schweizer Skiverband ‚Swiss Ski‘ auch 2008 noch die Bezeichnungen *Alpin Damen/Alpin Herren* etc. (vgl. <http://www.swiss-ski.ch/>, 29. 2. 2008), während der ‚Schweizerische Handball-Verband‘ von *ANM Männer* bzw. *ANM Frauen* etc. spricht (vgl. <http://www.handball.ch>, 29. 2. 2008).



Die *Dame* gehört im Schach zum spezifischen Fachvokabular; die Verwendung ist deshalb nicht weiter erstaunlich. In einem weiteren Beleg, der im Kontext der Sportart ‚Fechten‘ angelegt ist, sind nicht nur *Damen* sondern auch *Herren* erwähnt:

- (84) Leipzig ist aber auch für andere deutsche Fechter die Triebfeder. Die enttäuschenden Ergebnisse in Athen, wo es nur zu Silber durch die Degen-**Damen** und zu Bronze für die Degen-**Herren** gereicht hatte, sind kein Thema mehr. *Neue Zürcher Zeitung vom 29. April 2005, Ressort ‚Sport‘, wog: „Der Aufwärtstrend hält an. Die deutschen Fechter wollen spätestens an den WM in Leipzig wieder Weltklasse sein“.*

Ob in diesem Fall die Alliteration der *Degen-Damen* zu dieser Formulierung führte, oder ob ein anderer Grund dahinter steckt, kann nicht eruiert werden. Doch gibt es im ganzen Artikel keine weiteren Belege weder für *Frauen* noch *Damen*, weder *Männer* noch *Herren*.

Die Abkehr vom ‚Damensport‘ hin zum ‚Frauensport‘ gilt jedoch nicht für alle deutschsprachigen Medien. So zeigt eine Analyse der Sportberichterstattung im ‚Mannheimer Morgen‘ im Zeitraum 1995–2006 im DeReKo IDS (o.J.), dass dort auch 2006 *Damen* anstelle von *Frauen* verwendet wird.<sup>49</sup> Doch parallel dazu gibt es auch bereits 1995 Nennungen von *Frauen*.<sup>50</sup> Neben Sportlerinnen sind damit aber auch Frauen in anderen als Sportlerinnen-Rollen gemeint, d. h. als demographische Kategorie im Sinne von:

- (85) Frauen mögen lieber Boxen als Fußball *M95/505.00785 Mannheimer Morgen vom 31. Mai 1995, Ressort ‚FERNSEHEN‘: „Frauen mögen lieber Boxen als Fußball“.*

Der Sprachgebrauch im ‚Mannheimer Morgen‘ ist bezüglich des Gebrauchs von *Frauen* vs. *Damen* bzw. *Männer* vs. *Herren* in der Sportberichterstattung weniger systematisch als in der NZZ. Die Verwendung von *Damen* scheint im ‚Mannheimer Morgen‘ auch heute noch vorzuherrschen.

<sup>49</sup> Für die Recherche wurde ein ‚virtuelles Korpus‘ definiert durch die Anfrage: (Mannheimer /+w1 Morgen) UND Sport. Dieses Korpus umfasst 134 788 Texte mit 40 089 300 Wörtern. Darin wurde die Suchanfrage *Damen* ausgeführt, was 14 588 Treffer erzielte.

<sup>50</sup> Die Suchanfrage *Frauen* erzielte 8220 Treffer.

### 13.3.9 Ressorts im synchronen Vergleich: Sprachgebrauch in Leserbriefen

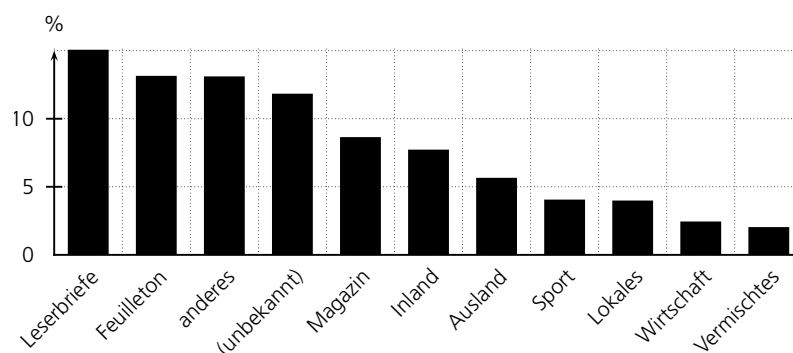
Die vorangehenden Beispielanalysen beruhen auf Typikprofilen, die auf der zeitlichen Achse definiert sind. D. h., die jeweils typischen Mehrworteinheiten sind innerhalb eines Ressorts typisch für eine Zeitperiode im Vergleich zu einer anderen Zeitperiode. Bei einer solchen diachronen Sichtweise ist nicht zu erwarten, dass primär Sprachgebrauchsmuster gefunden werden können, die typisch für eine bestimmte Textsorte oder pragmatische Funktion sind.

Ich habe aber auch für jedes Ressort ein Typikprofil erstellt, das die signifikantesten Mehrworteinheiten dieses Ressorts im Vergleich zu allen anderen Ressorts enthält. Diese Typikprofile sind online einsehbar.<sup>51</sup> Als Beispiel sei an dieser Stelle nur auf das Typikprofil des Ressorts ‚Leserbriefe‘ verwiesen, das in Tabelle A.11 auf Seite 359 aufgeführt ist. Zu den signifikantesten Mehrworteinheiten gehören Ausdrücke, die mit *Es ist* ... am Satzanfang beginnen. In Leserbriefen werden damit oft normative, deontische Aussagen eingeleitet:

- (86) **Es ist** unsere staatsbürgerliche Pflicht, am 18. April an die Urne zu gehen. Und die letzten Zweifel, wie er zu stimmen habe, wurden durch die eindruckliche Sendung beim Schreibenden ausgeräumt. *Neue Zürcher Zeitung vom 9. April 1999, Ressort ‚Briefe an die NZZ‘: „Sorgfältig ausgearbeitete Verfassung“*
- (87) Wenn ein Parlament volle sechs Jahre an einer Vorlage arbeitet, so muss doch irgendwo ein Wurm stecken. Der dürfte heissen: „Wie sag ich’s meinem Volk?“ **Es ist** noch lange nicht gut, was in einigen intellektuellen Köpfen steckt, es muss auch vom einfachen Bürger verstanden werden. *Neue Zürcher Zeitung vom 29. Juni 2001, Ressort ‚Briefe an die NZZ‘: „Am Volk vorbei?“*

Solche Aussagen sind typisch für bestimmte Textsorten in einer Zeitung, so erstaunt es nicht, dass die Mehrworteinheit *Es ist* signifikant häufiger in den Leserbriefen und im Feuilleton vorkommen, sowie in den Restgruppen ‚anderes‘ und ‚unbekannt‘, wie Abbildung 13.15 auf der nächsten Seite und Tabelle 13.32 auf der nächsten Seite zeigen. Der

<sup>51</sup> Vgl. <http://www.bubenhofer.com/korpusanalyse/>.



**Abbildung 13.15:** Die Frequenzen in den Ressorts des NZZ-Korpus von Artikeln die mindestens ein Vorkommen von *Es ist* aufweisen. Vgl. Tabelle 13.32 für die detaillierten Daten.

Ressort	<i>Es ist...</i>	
	#	%
<i>Ausland</i>	321	5,66
<i>Feuilleton</i>	729	13,15
<i>Inland</i>	377	7,73
<i>Leserbriefe</i>	154	15,07
<i>Lokales</i>	229	3,99
<i>Magazin</i>	255	8,65
<i>Sport</i>	209	4,06
<i>Vermischtes</i>	62	2,04
<i>Wirtschaft</i>	261	2,45
<i>(unbekannt)</i>	9	11,84
<i>anderes</i>	16	13,11
Total	2622	
$\chi^2 = 1154,889$ , $df = 10$		
$p < 0,001$ (signifikant)		

**Tabelle 13.32:** Verteilung und  $\chi^2$ -Statistik (Einheit: Artikel) von *Es ist* über die Ressorts des NZZ-Korpus. Absolute Frequenzen und am jeweiligen Ressort prozentuale Anteile der Artikel. Vgl. Abbildung 13.15 für eine grafische Darstellung der Verläufe.

Ausdruck *Es ist* kommt über alle Ressorts gesehen in 8 % aller Artikel vor, für ‚Leserbriefe‘ liegt der Wert bei 15 %, im ‚Feuilleton‘ bei 13 %.

Ebenso auffällig ist die Verwendung des Personalpronomens *ich* in den Leserbriefen. Das Typikprofil enthält 93 unterschiedli-

che Mehrworteinheiten, die *ich* enthalten, darunter Ausdrücke wie *als ... habe ich, ich ... Ablehnung der oder kann ich nur [sagen/begrüssen/festhalten/hoffen]*. Das Muster ALS [SELBSTBEZEICHNUNG] HABE ICH dient oft als Indikator für einen argumentativen Schritt:

- (88) Die Einseitigkeit des deutsch-tschechischen Vertrages ist offensichtlich. **Als** Schweizer sudetendeutscher Herkunft **habe ich** im Dezember 1996 bei Bundeskanzler Kohl, Aussenminister Kinkel und weiteren Politikern dagegen protestiert, dass die deutsche Bundesregierung für die Sudetendeutschen entscheidet und Verträge unterschreibt. *Neue Zürcher Zeitung vom 18. Februar 1997, Ressort ‚Briefe an die NZZ‘: „Bonn kann nicht für alle Sudetendeutschen sprechen“*
- (89) **Als** jährlicher Mithelfer am 1. Mai 1996 an einem der Essenstände **habe ich** als Unbeteiligter die Eskalation der „Nachdemo“ ziemlich zu spüren bekommen. *Neue Zürcher Zeitung vom 10. Mai 1996, Ressort ‚Briefe an die NZZ‘: „Die Krawalle am 1. Mai in Zürich“*

Bei beiden Belegen dient die Erwähnung der eigenen Person bzw. Funktion als ARGUMENTUM AD HOMINEM (Kienpointner 1992, 321); die eigene Erfahrung wird als Argument dafür verwendet, dass man eine überzeugende Aussage machen kann.

Im Typikprofil der Leserbriefe finden sich aber auch parolenartige Ausdrücke, so z. B. *ohne Bauern stirbt die Stadt* oder *[das ist] der Preis der Naivität*, oder Indikatoren für argumentative Zusammenhänge wie *nichts zu tun* oder *[stellt] sich die Frage*.

Die Leserbriefe in der NZZ unterscheiden sich wahrscheinlich von Leserbriefen in anderen Zeitungen und Medien. Es wäre interessant, bezüglich der verwendeten Mehrworteinheiten Vergleiche herzustellen. Ich werde in Kapitel 17.1 kurz auf die Analyse eines Web-Diskussionsforums eingehen, um das Potenzial von solchen synchron ausgerichteten Spachgebrauchsvergleichen aufzuzeigen.



## 14 Hypothesenbildung

Das vorangehende Kapitel 13 brachte eine Reihe von Einzelbeobachtungen hervor, die für sich gesehen interessant sind, jedoch noch ein eher heterogenes Feld bilden. Grundsätzlich lässt sich diese Heterogenität in der vorliegenden Arbeit nicht auflösen, wie ich bereits zu Beginn dieses Teils (vgl. Seite 187) betont habe: Die Analysen haben illustrativen Charakter und die Heterogenität ist das Resultat des corpus-driven-Zugangs. Die corpus-driven entstandenen Daten wurden in der vorliegenden Arbeit nur zu einem Bruchteil ausgewertet. Würden die Analysearbeiten weitergeführt und auch um diskurslinguistische Standardverfahren ergänzt, könnten die im Folgenden genannten Hypothesen genauer überprüft werden. Im besten Fall wäre es möglich, sie weit besser zu fundieren. Oder aber es würde sich zeigen, dass die Hypothesen nicht haltbar sind oder angepasst werden müssen. In jedem Fall entstünden aber bei der weiteren Ausschöpfung des Materials zusätzliche Hypothesen zum Sprachgebrauch im NZZ-Korpus und weitere – und vor allem viel detailliertere – Beschreibungen von Diskursen.

Trotz der Heterogenität der Analysen in der vorliegenden Arbeit stehen jeweils ähnliche Fragestellungen im Hintergrund. Sie hatten die Auswahl der Mehrworteinheiten beeinflusst, die genauer analysiert wurden. Im Vordergrund steht die Frage, wie sich Sprechweisen im NZZ-Korpus diachron verändern und wie sie sich synchron unterscheiden. Die für jeweils bestimmte Zeitperioden oder Textfunktionen (meist: Ressorts) typischen Mehrworteinheiten konnten teilweise zu abstrakteren Mustern erweitert werden.

Die Heuristik der korpuslinguistischen Diskursanalyse, die ich in Kapitel 5.2 ausgeführt habe (vgl. auch Abbildung 5.1 auf Seite 104), fordert nun eine Beschreibung von Diskursen basierend auf den Interpretationen dieser Muster. Es ergaben sich folgende Hinweise auf Diskurse.

### 14.1 Sprechen über KRIEG UND GEWALT

Die Redeweisen in kriegesischen Kontexten weisen einerseits eher kontextunabhängige und andererseits kontextabhängige Muster auf. Durch die corpus-driven-Analyse des NZZ-Korpus sind besonders die folgenden Muster aufgefallen:

- I. TYPISCHE SPRACHGEBRAUCHSMUSTER: Die Zeitperioden 1995–1997 und 2003–2005 weisen im Ressort ‚Ausland‘ je spezifische Sprachgebrauchsmuster auf, die auf je typische Protagonisten, Themen und Ereignisse verweisen.

Für 1995–1997 sind das:

- a) Die Jugoslawienkriege, wobei zu den typischsten Muster *die bosnischen Serben, der in Bosnien, in der Serben, der Nato und und der in Sarajewo* gehören.
- b) Die Friedensnobelpreisträgerin und Oppositionspolitikerin *Aung San Suu Kyi* in Burma war zu dieser Zeit immer wieder Thema.
- c) Zu den weiteren Akteuren gehören der südkoreanische Oppositionspolitiker *Kim Dae Jung*, der damalige *Bundeskanzler Kohl*, *US-Präsident Clinton* und der *russische Präsident Jelzin*.

Für 2003–2005 sind das:

- a) Der Krieg *im Irak* wird mit den Akteuren *Präsident Bush*, *Usama bin Ladin* und *Saddam Hussein* erwähnt. Die wichtigsten Themen sind der *Abzug aus dem Irak* und die *Zahl der Opfer*.
- b) Der *Abzug aus dem Gazastreifen* und die *besetzten Gebiete* spielen eine große Rolle; als Akteur *Ministerpräsident Sharon*.

Die typischsten Sprachgebrauchsmuster werden mehrheitlich in Kontexten von Krieg und Gewalt verwendet, mit Ausnahme von Namen von Spitzenpolitikern, deren Kontexte mehrdeutiger sind.<sup>1</sup>

<sup>1</sup> Vgl. Kapitel 13.1.1.

2. ETHNIENBEZEICHNUNGEN verweisen auf problematische, kriegsrische Kontexte, LÄNDERBEZEICHNUNGEN eher auf Kontextualisierungen von Normalität und wirtschaftlichem Aufschwung.<sup>2</sup>
3. KAMPFFLOSKELN wie KAMPF GEGEN X und KAMPF DEM X erfüllen je spezifische pragmatische Funktionen der Kontextualisierung von Kampf. Während erstere eher beschreibend verwendet werden, weist letztere Parolen-Charakter auf und transportiert zudem dazugehörige Meinungen. In Zeitungstexten ist diese Parole deshalb typisch für argumentative Passagen in Leserbriefen. Besonders ausgeprägt ist diese Unterscheidung in den Mustern KAMPF GEGEN (DEN) TERROR bzw. KAMPF DEM TERROR nach dem 11. September 2001. Im deutschsprachigen Raum gibt es inzwischen einige Belege für *Kampf dem Terror*. Die Verwendung dieses Sprachgebrauchsmusters ist ein Indikator für Kritik an diesem ‚Kampf‘. Trotzdem ist *Kampf gegen (den) Terror* weitaus gebräuchlicher. Die Verwendung dieser KAMPFFLOSKELN mit *Terror* sind ein wichtiger Indikator für den Status, den die Terrorbekämpfung hat. Die vermehrte Verwendung von *Kampf dem Terror* würde z. B. auf einen veränderten Status von *Terror* schließen lassen, einen ähnlichen Status, wie er heute beispielsweise für *Krebs* oder *Stau* (*Kampf dem Krebs/Stau*) gilt.<sup>3</sup>
4. NENNUNGEN VON ZÄHLENSWERTEM, vor allem in der Form ZAHL DER X, verweisen auf Kontexte, die als problematisch erkannt werden. Die Kontexte haben sich in der Zeit nach 2000 deutlich in Richtung Kriminalität und Krieg bewegt, während sie vor dem Jahr 2000 noch eher politisch waren.<sup>4</sup>
5. Die Art des REDENS ÜBER DAS STERBEN ist ein weiterer Indikator für die Gestalt von Diskursen im Bereich von Krieg und Gewalt. Im Ressort ‚Ausland‘ hat die Verwendung des Paradigmas STERBEN seit 1998 und vor allem mit dem Jahr 2001 zugenommen. Die potenziellen Verwendungskontexte haben sich verschoben: Sie sind vermehrt in Ausland-Themen zu finden. Es

---

<sup>2</sup> Vgl. Kapitel 13.1.2.

<sup>3</sup> Vgl. Kapitel 13.1.3.

<sup>4</sup> Vgl. Kapitel 13.1.4.



sind vor allem die Ausdrucksvarianten *(ge)tötet* und *in den Tod*, die seit 2001 zugenommen haben. Die Beobachtungen geben Hinweise darauf, dass im internationalen, politischen Kontext vermehrt von tödlichen Ereignissen die Rede ist. Zudem hat sich die Sprechweise leicht verändert: Die Ausdrucksvariante *in den Tod* (*gerissen*) bzw. (*riss*) *in den Tod* hat signifikant zugenommen.<sup>5</sup>

## 14.2 Sprechen über SORGEN UND PROBLEME

Es liegt in der Natur des Mediengeschäfts über Probleme, Krisen und Brennpunkte zu berichten. Deshalb erstaunt es nicht, dass auch im NZZ-Korpus diejenigen politischen, kriegerischen und gesellschaftlichen Problemthemen gefunden werden, die auch in anderen Medien oder im gesellschaftlichen Diskurs vorherrschen.

Aus korpuslinguistischer Sicht interessiert jedoch, wie und wo genau diese Probleme im Korpus thematisiert werden. Eine Auflistung aller im Korpus vorkommenden Nomina, geordnet nach Frequenz, würde diese Themen nicht widerspiegeln. Es handelt sich um bestimmte syntagmatische Muster, die die Slots für diese Themen bilden.

Aus diskurslinguistischer Sicht überrascht die Tatsache, wie einfach und beinahe erschöpfend über solche syntagmatischen Muster Themen gefunden werden, die z. B. in einer demoskopischen Untersuchung wie dem ‚Sorgenbarometer‘<sup>6</sup> von der Bevölkerung ebenfalls als die ‚Top Ten‘ der Probleme angesehen werden. Auch ergeben sich Parallelen zu einer anderen Textsorte: Den Verhandlungen im nationalen Parlament.

Die folgenden Sprachgebrauchsmuster markieren im NZZ-Korpus Problemthemen des politischen und gesellschaftlichen Diskurses:

- I. KAMPFRHETORIK: Die Muster BEKÄMPFUNG VON X/KAMPF GEGEN X bringen Entitäten hervor, die wörtlich oder zumindest semantisch mit den meistgenannten Problemen des ‚Sorgenbarometers‘ der Schweiz übereinstimmen. Weitere Übereinstim-

<sup>5</sup> Vgl. Kapitel 8.2.3.

<sup>6</sup> Vgl. Kapitel 13.2.3.

mungen ergeben sich, wenn im ‚Amtlichen Bulletin‘, den Wortprotokollen des schweizerischen Parlaments, nach ebendiesen Mustern gesucht wird. Es sind nur wenige Themen, die aus syntaktischen oder semantischen Gründen nicht über dieses Muster gefasst werden können – oder umgekehrt argumentiert: In der medialen Wirklichkeit wird das als wichtiges Problem angesehen, wogegen *ein Kampf* geführt oder das *bekämpft* werden kann.<sup>7</sup>

Die Analyse der Muster KAMPF GEGEN X, KAMPF DEM X und KAMPF MIT X im Ressort ‚Ausland‘ zeigte darüber hinaus systematische Gebrauchsdifferenzierungen.<sup>8</sup> Die drei Muster weisen zwar alle denselben propositionalen Gehalt auf, zeigen aber deutliche semantische Präferenzen der Füllungen für den Slot X. Solche Differenzierungen können in die Analyse von Sorgen und Problemen einbezogen werden.

2. ARGUMENTATIONSFLOSKELN: In argumentativen Kontexten hat das Sprachgebrauchsmuster NICHT NUR X SONDERN AUCH Y gezeigt, wie die Füllungen der Slots X und Y in zeitlicher Abhängigkeit stehen. Oft wird mit der Nennung eines Sachverhalts in der zweiten Position (Y) diesem den Status der Unterschätzung zugewiesen: Der zweite Sachverhalt, nicht der erste, sei wirklich von Bedeutung.<sup>9</sup>

Damit gehört dieses Sprachgebrauchsmuster zu den Gradmessern, die anzeigen, welche Sachverhalte in einem Diskurs als problematisch eingeschätzt werden, oder von denen behauptet wird, dass ihre Problematik im Diskurs (noch) unterschätzt wird.

Es ist klar, dass neben diesem Muster noch einige mehr existieren, die solche Funktionen aufweisen, z. B. das Muster X IST NICHT MEHR Y.<sup>10</sup> Es ist auch deutlich, dass diese Funktionen oft nur für bestimmte Textsorten gelten, z. B. weist das oben

<sup>7</sup> Vgl. Kapitel 13.2.3.

<sup>8</sup> Vgl. Kapitel 13.1.3.

<sup>9</sup> Vgl. Kapitel 13.2.4.

<sup>10</sup> Vgl. Kapitel 13.3.2.

genannte Muster hauptsächlich in argumentativen Kontexten diese Funktionen auf.

3. NENNUNG VON ZÄHLENSWERTEM: Wie bereits für Diskurse im Bereich KRIEG UND GEWALT (vgl. Kapitel 14.1) gilt auch im Kontext von SORGEN UND PROBLEME, dass Gezähltes ein Indikator für problematische Themen ist. In der vorliegenden Untersuchung bin ich zwar primär auf die Muster ZAHL DER X bzw. [ZAHL] X im Ressort ‚Ausland‘ eingegangen, doch gelten ähnliche Befunde auch für das Inland-Ressort.<sup>11</sup>

### 14.3 Sprachgebrauch und Welt

Die zeitliche Gebundenheit von vielen sprachlichen Mustern sticht hervor, wie jeweils die Berechnung typischer Mehrworteinheiten zweier Perioden gezeigt hat.<sup>12</sup> Diese sprachlichen Veränderungen lassen sich damit durch textexterne Faktoren erklären. Die Frage der Kausalität jedoch bleibt offen: Sind es Ereignisse in der Welt, die ihren Niederschlag im NZZ-Korpus finden, oder ist es zuerst die sprachliche Veränderung im NZZ-Korpus, die sich auf eine andere Sicht auf Welt auswirkt? So hat die Analyse der Lemmata *Damen* vs. *Frauen* bzw. *Herrn* vs. *Damen* in der Sportberichterstattung deutlich gemacht, dass sich die Ausdrucksweisen (z. B. *Damen-Tennis* vs. *Frauen-Tennis*) in der NZZ schlagartig geändert haben. Dafür wird wahrscheinlich eine redaktionelle Direktive verantwortlich sein, doch warum hat sich die Redaktion dazu entschlossen? Die gesellschaftlichen Faktoren dafür bleiben diffus, der neue Sprachgebrauch wird aber eine andere Sicht auf Welt auslösen.

Hingegen ist klar, dass der textexterne Faktor ‚Anschläge vom 11. September 2001‘ zu sprachlichen Veränderungen geführt hat. Doch wie der Kampf gegen Terrorismus geführt wird, ob als *Kampf gegen den Terror*, *Kampf mit dem Terror* oder mit Parolen als *Kampf dem Terror!* umschrieben wird, hat Auswirkungen auf die Art, wie dieser Kampf geführt wird. Dies betont beispielsweise die 2008 erschienene Studie „How terrorist groups end: lessons for countering Al Qa’ida“

<sup>11</sup> Vgl. Kapitel 13.1.4.

<sup>12</sup> Vgl. die Kapitel 13.1.1, 13.2.1 auf Seite 241 und 13.3 auf Seite 268.

(Jones/Libicki 2008), die dafür plädiert, die Bezeichnung *war on terrorism* aufzugeben:

This would include ending the notion of a „war“ on terrorism and replacing it with such concepts as *counterterrorism*, which most governments with a significant terrorism problem use. This change might seem pedantic but would have significant symbolic importance. Moving away from military references would indicate that there was no battlefield solution to countering terrorism. (Jones/Libicki 2008, 125)

In der vorliegenden Untersuchung sind folgende Sprachgebrauchsmuster zutage getreten, die eine (wie auch immer gerichtete) kausale Verbindung von Sprache und Welt repräsentieren:

1. ORTE UND PERSONEN: Die Nennung von Orten und Personen ist deutlich zeitabhängig, wie vor allem die typischen Mehrwort-einheiten für 1995–1997 bzw. 2003–2005 im Ressort ‚Ausland‘ gezeigt haben.<sup>13</sup> Das gilt generell auch für Nationen und Nationalitätenbezeichnungen und beschränkt sich nicht auf das Auslands-Ressort.<sup>14</sup> Im Ressort ‚Wirtschaft‘ zeigten sich darüber hinaus Nennungen von Institutionen (Börsen, Finanzinstitutionen) als zeitabhängig.<sup>15</sup>
2. KONTEXTUALISIERUNGSPROFILE: Ganz generell zeigen die meisten entdeckten Sprachgebrauchsmuster zeitabhängige Kontextualisierungsprofile.
  - a) Es existieren einerseits Muster, deren Slots zeittypische Füllungen aufweisen. Dazu gehören (zum Teil bereits oben genannte) Muster wie KAMPF GEGEN X, DIE DEUTSCHE(N) X oder DIE [ADJEKTIV] GESELLSCHAFT.<sup>16</sup>
  - b) Andererseits gibt es Sprachgebrauchsmuster, die für bestimmte zeitliche Perioden typisch sind. So z. B. die Muster des REDENS ÜBER STERBEN, die semantischen Präferenzen der KAMPFRHETORIK, die ANGLIZISMEN IN INSTITUTIONENNAMEN, die GESCHLECHTERBEZEICHNUNGEN *Damen*

<sup>13</sup> Vgl. Kapitel 13.1.1.

<sup>14</sup> Vgl. Kapitel 13.1.2 und 13.3.5 auf Seite 274.

<sup>15</sup> Vgl. Kapitel 13.3.4.

<sup>16</sup> Vgl. zu KAMPF GEGEN X Kapitel 13.2.3, für DIE DEUTSCHE(N) X Kapitel 13.3.6 und DIE [ADJEKTIV] GESELLSCHAFT Kapitel 13.3.1.

vs. *Frauen* bzw. *Herren* vs. *Damen* in der Sportberichterstattung oder MEDIALE METABESCHREIBUNGEN wie *vor den Medien*.<sup>17</sup>

#### 14.4 Sprachgebrauch und Text

Mit dieser vierten Kategorisierung möchte ich die diskursive Gebundenheit in pragmatisch-funktional definierten Texten verdeutlichen. Aus pragmatischer Perspektive weisen Texte bestimmte Funktionen auf und lassen sich nach diesem Kriterium z. B. nach Textsorten unterscheiden.

Praktisch alle Beispielanalysen der vorliegenden Arbeit beruhen auf Typikprofilen, die auf der zeitlichen Achse definiert sind, also die jeweils typischen Mehrworteinheiten aufführen, die innerhalb des Ressorts im Vergleich zwischen zwei Zeitperioden signifikant sind. Deshalb sind bei den Analysen auch primär Sprachgebrauchsmuster, die zeittypisch sind, aufgefallen und beschrieben worden. Trotzdem gab es bei einigen Mehrworteinheiten Hinweise darauf, dass sie weniger zeittypisch, sondern typisch für ganz bestimmte Textsorten oder pragmatische Funktionen innerhalb von Texten sind. Darunter fallen folgende Muster:

1. BERICHTEN VS. KOMMENTIEREN: Besonders die Analyse des Typikprofils von Leserbriefen (vgl. Kapitel 13.3.9) hat einige Sprachgebrauchsmuster hervorgebracht, die typisch für kommentierende oder argumentierende Textpassagen sind. Am auffälligsten ist die musterhafte Verwendung von *Es ist...* an Satzanfängen, das besonders oft in Leserbriefen und im Feuilleton-Ressort verwendet wird und normative, deontische Aussagen einleitet. Weiter gibt es viele Mehrworteinheiten mit dem Personalpronomen *ich*, die in den anderen Ressorts unüblich sind. Der Ausdruck *als... habe ich* kann beispielsweise als ARGUMENTUM AD HOMINEM verwendet werden.

<sup>17</sup> Vgl. zu REDEN ÜBER STERBEN Kapitel 8.2.3, zur KAMPFRHETORIK Kapitel 13.1.3, für die ANGLIZISMEN IN INSTITUTIONENNAMEN Kapitel 13.3.7, zu den GESCHLECHTSBEZEICHNUNGEN Kapitel 13.3.8 auf Seite 287 und zu MEDIALEN METABESCHREIBUNGEN Kapitel 13.2.2.

Bei der Analyse der Kontextualisierungsprofile von *Kampf gegen X*, *Kampf dem X* und *Kampf mit X* haben sich deutliche Differenzen im Gebrauch gezeigt. Dabei handelt es sich nicht nur um semantische Präferenzen (vgl. dazu Kapitel 14.2), sondern um pragmatisch unterschiedliche Funktionen. So wird der *Kampf gegen den Terror* im Ausdruck *Kampf dem Terror* zur Parole; der Ausdruck findet sich besonders häufig in argumentativen Kontexten und zeigt eine Distanzierung oder Kritik an.<sup>18</sup>

2. EINORDNEN UND BEWERTEN: Mit dem Muster NICHT NUR X SONDERN AUCH Y werden in argumentativen Kontexten zwei Sachverhalte miteinander in eine hierarchische Beziehung gesetzt. Oft wird der zweite Sachverhalt Y dem ersten übergeordnet und damit gleichzeitig behauptet, er werde im Vergleich zum ersten Sachverhalt unterschätzt (vgl. auch Kapitel 14.2).<sup>19</sup>

Auch die Muster X IST NICHT MEHR [= *weiterhin*] Y bzw. X IST NICHT MEHR [= QUANTITATIV] Y zeigen Tendenzen zu solchen pragmatischen Funktionen, mit denen Sachverhalte eingeordnet oder bewertet werden, wobei die Untersuchungsbasis zu wenige Belege ergab, um diese Hypothese statistisch beweisen zu können.<sup>20</sup>

3. TEXTSORTENSPEZIFIKA: Eine Textsorte kann als Bündel bestimmter pragmatischer Funktionen aufgefasst werden. Deshalb ist nicht erstaunlich, dass sich eine Vielzahl von Mustern der Textoberfläche als typisch für bestimmte Textsorten erweist. Die oben erwähnten Funktionen EINORDNEN UND BEWERTEN bzw. KOMMENTIEREN sind im NZZ-Korpus typischerweise in Leserbriefen oder Kommentaren erfüllt.

Darüber hinaus finden sich auch Sprachgebrauchsmuster, die sogar nur für einen Teilbereich eines Ressorts, der ‚Wirtschaft‘, typisch sind: Die Börsenberichterstattung. Dazu gehören Ausdrücke wie *mit Plus von*, *Yen in Franken*, *im dritten Quartal*, *tiefer als am* etc. Auch für andere Textsorten gibt es solche typi-

<sup>18</sup> Vgl. Kapitel 13.1.3.

<sup>19</sup> Vgl. Kapitel 13.2.4.

<sup>20</sup> Vgl. Kapitel 13.3.2.

schen Mehrworteinheiten, wie die Grobanalysen der anderen Ressorts gezeigt haben.<sup>21</sup>

---

<sup>21</sup> Vgl. zur Börsenberichterstattung Kapitel 13.3 und 13.3.4 auf Seite 273, für die anderen Ressorts Kapitel 13.1.1 und 13.2.1 auf Seite 241.

## **IV**

### **Fazit und Ausblick**





Im letzten Teil der vorliegenden Arbeit möchte ich auf vier Aspekte zu sprechen kommen:

1. Basierend auf theoretischen Überlegungen zu einer Diskurs- und Kulturanalyse (Teil I) sah ich im Konzept der Sprachgebrauchsmuster den Ausgangspunkt für die weiteren Überlegungen. War das der richtige Ausgangspunkt (Kapitel 15.1)?
2. Darauf aufbauend haben sich sieben linguistische Konzepte angeboten, um Sprechweisen in Diskursen zu operationalisieren (Kapitel 4). Bereits bei der Herleitung dieser Konzepte hat sich gezeigt, dass sich nicht alle für eine Analyse der Textoberfläche eignen würden. Waren die Einschätzungen richtig (Kapitel 15.2)?
3. Mit dem Grundsatz einer corpus-driven-Analyse von Textdaten habe ich in Teil II eine Heuristik entwickelt, die theoretischen Vorüberlegungen entgegen kommen sollte, gleichzeitig aber auf ebendiese theoretischen Überlegungen zurückwirkt. Inwiefern geschieht das tatsächlich und mit welchen Folgen (Kapitel 16.1)?
4. Schließlich stellt sich schlicht die Frage, ob die vorgeschlagene Heuristik einer korpuslinguistischen Diskursanalyse überhaupt gewinnbringend zur Beantwortung von diskurs- und kulturanalytischen Fragestellungen dient (Kapitel 16.2).

Im Anschluss an die Evaluation der geleisteten Arbeit stelle ich in einem Ausblick Überlegungen an, in welche Richtung eine korpuslinguistische Diskurs- und Kulturanalyse weiterentwickelt werden müsste (Kapitel 17).



## 15 Theorie und Praxis

### 15.1 Sprachgebrauchsmuster als Kristallisationkerne von Diskursen

Musterhafter Sprachgebrauch ist das Resultat von kooperativem sprachlichen Handeln. Bestimmte Sprachgebrauchsmuster zeigen deshalb bestimmte Aspekte sprachlichen Handelns an und können so als charakteristische Eckpfeiler von Diskursen gelesen werden. Mit dieser These begründete ich in Kapitel 3 den eingeschlagenen Weg, die Sprachgebrauchsmuster als Indikatoren für Diskurse zu lesen. Musterhaft verwendete sprachliche Einheiten, die die Wortgrenze überschreiten, bilden also „minimale Kristallisationkerne“ (Linke 2003a, 40) von Diskursen.

Während etwa in begriffsgeschichtlicher Tradition der Fokus auf Wörtern liegt, versuchte ich diese ‚Kristallisationkerne‘ weiter zu fassen: Erstens wird das Schwergewicht nicht auf Einzelwörter, sondern auf Wortverbindungen gelegt, die in einem Korpus überzufällig häufig zusammen auftreten. Dahinter steckt die Idee, die ‚Typik des Sprechens‘ zu erfassen, indem nach ‚idiomatischen Prägungen‘ im Sinne von Feilke (1993) (vgl. Kapitel 4.4) gesucht wird.

Zweitens versuchte ich das rekurrente Auftreten (die musterhafte Verwendung dieser Wortverbindungen) als ‚Sprachgebrauchsmuster‘ zu lesen, die Indikatoren für die Gestalt von Diskursen darstellen. Denn im Anschluss an Foucault (1981) formen die Prozeduren der Produktion des Diskurses das inhaltlich Sagbare, sowie die Sprechweisen, also die diskursive Praxis. Diese diskursive Praxis ist nichts anderes als die regelmäßige, eben: musterhafte Art des Aussagens (vgl. Kapitel 2.3).

Sprachgebrauchsmuster sind ein Phänomen auf der Textoberfläche. Das ist für eine korpuslinguistische Diskursanalyse wichtig, da es die maschinelle, quantitative Verarbeitung von Texten vereinfacht. Damit ist die Textoberfläche Ausgangspunkt für die weitere Analyse.

Dieser Ausgangspunkt steht im Einklang mit der Rehabilitierung der sprachlichen Oberfläche, wie sie die „zweite pragmatische Wende“ (Feilke 2003, 217) postuliert (vgl. Kapitel 3.1.1). Nicht semantische Tiefenstrukturen müssen entschlüsselt werden, um den pragmatischen Wert von Texten zu bestimmen, sondern die Formulierungen auf der Textoberfläche sind es, die Ausdruck von Sprachgebrauchswissen sind. Diese Formulierungen schaffen die übereinstimmenden Kontexte für unterschiedliche Leserinnen und Hörer und erlauben so gegenseitiges Verstehen.

Wie decken sich diese theoretischen Überlegungen mit den korpuslinguistischen Befunden? Es ist gut möglich, Sprachgebrauchsmuster als statistisch fassbares Phänomen zu operationalisieren. Die in der vorliegenden Arbeit verwendeten Methoden sind vergleichsweise simpel. Ich habe mit statistischen Signifikanztests gearbeitet, um Wortverbindungen aus den Daten zu extrahieren und ihre Typik für bestimmte Teilkorpora zu bestimmen (vgl. Kapitel 7 und 12). Ich ging dabei von unbearbeitetem Wortmaterial im Korpus aus, und nicht etwa von lemmatisierten oder mit Wortarten annotierten Daten, was zudem den technischen Aufwand weiter reduzierte, aber auch theoretisch sinnvoll ist. Denn die Analysen haben gezeigt, dass in den unterschiedlichen syntaktischen Ausprägungen der Wortformen entscheidende Bedeutungsunterschiede liegen können. Dieser Befund wird durch ähnliche Einschätzungen von anderen korpus- und computerlinguistischen Arbeiten gestützt (vgl. Kapitel 6.5) und kann auch theoretisch begründet werden, wie die Ausführungen zu den idiomatischen Prägungen (Kapitel 3.1.1) gezeigt haben.

Die corpus-driven extrahierten Mehrworteinheiten sind noch keine Sprachgebrauchsmuster, sondern zunächst bloß statistisch auffällige und für bestimmte Teilkorpora typische Ausdrücke. Erst weitere corpus-based Analysen, mit denen die Mehrworteinheiten (in Abhängigkeit des Forschungsinteresses) kategorisiert, ggf. zusammengefasst oder erweitert werden, führen zu Sprachgebrauchsmustern.

Auch wenn mit dem corpus-driven-Zugang versucht wird, möglichst ohne theoretische Vorannahmen die Daten zu bearbeiten, müssen trotzdem eine Reihe von Parametern für die Berechnungen gesetzt werden. So muss festgelegt werden, wie viele Wörter eine Mehrworteinheit haben darf, wie weit diese Wörter auseinanderliegen dürfen etc.

(vgl. Kapitel 6.1). Und primär wurde entschieden, Sprachgebrauchsmuster als rekurrente Kombination von Wörtern zu definieren. Das ist zwar naheliegend, wenn die sprachliche Oberfläche als Ausgangspunkt für die Analysen dienen soll, aber auch die Textoberfläche böte mit Phonemen, Morphemen, typografischen Phänomenen, Kategorien der Textgliederung etc. alternative Elemente, die genutzt werden könnten.

Der induktive Zugang ist also nicht vorurteilslos, aber er definiert seine Grundannahmen anders als es z. B. semantisch-tiefenstrukturelle Analysen tun. Und dieser Zugang führt dann auch zu einem bunten Strauß an Sprachgebrauchsmustern, die einerseits eher Kristallisationspunkte des inhaltlich Sagbaren, andererseits der Sprechweisen, der diskursiven Praxis darstellen. Diese Heterogenität ist auf der Ebene der Analyse nicht immer einfach zu handhaben und es ist kaum möglich, daraus homogene Diskursbeschreibungen abzuleiten. Meine Analysen, die zwar nur exemplarischen Charakter haben, zeigten dies deutlich (vgl. Kapitel 14).

Doch legten die Analysen m. E. auch einige Sprachgebrauchsmuster frei, die eindeutig Indikatoren für diskursive Praktiken sind. In einem KRIEGSDISKURS wie im Kontext ‚Irakkrieg‘ scheint es beispielsweise Muster zu geben, die erstens typisch für genau diesen Diskurs sind und zweitens die Selektion aus einem Paradigma möglicher Muster darstellen, wobei die Selektion eben durch die Prozeduren der Produktion des Diskurses gesteuert ist. Die exemplarisch herausgearbeiteten Muster KAMPF GEGEN X vs. KAMPF DEM X, ETHNIENBEZEICHNUNGEN oder das Paradigma REDEN ÜBER DAS STERBEN mit diskursspezifischen Selektionspräferenzen entsprechen diesem Typus (vgl. Kapitel 14.1).

In einer Reihe von Mustern hat sich Sprachgebrauchswissen manifestiert. Es sind dies Muster, die textsortenspezifisch sind oder deutliche pragmatische Funktionen aufweisen und, weil diese Phänomene musterhaft sind, von unterschiedlichen Rezipienten anscheinend ähnlich kontextualisiert werden. Besonders die in Kapitel 14.4 aufgeführten Sprachgebrauchsmuster gehören in diese Klasse, so z. B. NICHT NUR X SONDERN AUCH Y oder textsortenspezifische Ausdrücke wie *im dritten Quartal*, *tiefer als am* etc.

Aufgrund der vorliegenden Befunde ist es nicht gewagt zu behaupten, dass Diskurse als Ensembles von Sprachgebrauchsmustern, also

Kristallisationskernen, konstruiert werden können. Die Frage aus korpuslinguistischer Sicht lautet eher: Wie kann erreicht werden, dass einerseits möglichst viele dieser Sprachgebrauchsmuster gefunden werden und andererseits die meisten der gefundenen Muster auch tatsächlich Kristallisationskerne sind? Oder in Termini der Informationsverarbeitung: Wie kann ein optimales Verhältnis zwischen ‚Precision‘ (Treffgenauigkeit) und ‚Recall‘ (Ausbeute) erreicht werden?

### 15.2 Semantische Konzepte in der korpuslinguistischen Operationalisierung

Eines der Ziele der vorliegenden Arbeit war, den Fokus auf die Sprechweisen in Diskursen zu legen. Dafür habe ich semantische Konzepte der Linguistik, die mir dafür geeignet schienen, beigezogen (vgl. Kapitel 4). Es handelt sich um die Konzepte ‚Stil‘, ‚kommunikative Gattungen‘, ‚Mentalitäten und diskurssemantische Grundfiguren‘, ‚Typik des Verhaltens und idiomatische Prägungen‘, ‚Argumentationsfiguren und Topoi‘, ‚Metaphern‘ und ‚Sprachgebrauchsgeschichte als Kulturanalyse‘. Diese Ansätze hätten die Suchmethode nach den Sprachgebrauchsmustern theoretisch begründen und die Einordnung der gefundenen Sprachgebrauchsmuster vereinfachen sollen. Die Beschäftigung mit den heterogenen Konzepten ließ vermuten, dass sich besonders das Konzept der idiomatischen Prägungen und der Ansatz der Sprachgebrauchsgeschichte für eine induktive, sich an der Textoberfläche orientierenden korpuslinguistischen Diskursanalyse eignen würden (vgl. Kapitel 4.8). Im Kapitel 15.1 oben habe ich bereits dargestellt, dass die beiden Ansätze die Begründung liefern, warum Sprachgebrauchsmuster als Kristallisationskerne von Diskursen gelesen werden können.

Für die restlichen Ansätze hat sich die eher pessimistische Einschätzung über die Nützlichkeit für mein Vorhaben weitgehend bestätigt. Und umgekehrt: Die korpuslinguistische Methode, wie ich sie in der vorliegenden Arbeit vorschlage, lässt sich nicht produktiv für die Fragestellungen dieser Ansätze verwenden. Zwar ist es ansatzweise möglich, einzelne Sprachgebrauchsmuster als Befunde für diese Ansätze zu begreifen. So kann die Selektion von *Frauenhandball* statt

*Damenhandball* als stilistische Selektion aufgefasst werden, mit der ein „Was“ durch ein „Wie“ im Hinblick auf ein „Wozu“ (Fix 1996, 310) ausgedrückt wird. Und die Verfahren, um die typische Sprechweise in einem Teilkorpus über die dafür typischen Sprachgebrauchsmuster zu bestimmen, deckt sich zum Teil mit der Vorstellung von ‚Stil‘: Stil zeigt sich im Unikalen eines Textes im Vergleich zum Überindividuellen. Allerdings hat sich der korpuslinguistische Ansatz einer Diskursanalyse gerade vom Individuum abgesetzt und vergleicht, wenn schon, überindividuelle Stile mit anderen überindividuellen Stilen, indem die typischen Sprechweisen eines Teilkorpus im Vergleich zu einem anderen Teilkorpus herausgearbeitet werden.

Einzelne Sprachgebrauchsmuster könnten als Bausteine von kommunikativen Gattungen gelesen werden, so z. B. die Ausdrücke der Börsenberichterstattung, die sehr formalisiert ist (vgl. Kapitel 13.3.4). Hilfreicher ist jedoch die theoretische Begründung der kommunikativen Gattungen im sozialen Handeln, die das Denkbare konstruieren und reproduzieren und damit das Sagbare beeinflussen (vgl. Kapitel 4.2). Die kommunikativen Gattungen bieten dadurch das Potenzial einer linguistischen Fundierung der Foucault’schen ‚diskursiven Praxis‘.

‚Mentalitäten‘ beschreiben das kollektive Fühlen, Denken und Wollen. Die Mentalitäten stecken die Grenzen des Sagbaren in einer bestimmten Gesellschaft zu einer bestimmten Zeit ab. Die Operationalisierung dieses Konzepts kann die Analysekatgorie der ‚diskurssemantischen Grundfigur‘ leisten, die semantische Systeme organisiert und strukturiert (vgl. Kapitel 4.3). Eine Korpusanalyse kann keine diskurssemantischen Grundfiguren offen legen, zumindest nicht unmittelbar. Dazu ist ein hermeneutischer Zugriff nötig, mit dem versucht wird, eine größere Menge an Sprachgebrauchsmustern als Indikator für solche Grundfiguren, z. B. für eine semantische Opposition, zu lesen. Meine Analysen sind zu wenig vollständig, um eine Grundlage für diesen Zugriff bieten zu können. Darüber hinaus müsste das Vorgehen hinterfragt werden, alleine mit Daten der Textoberfläche zu arbeiten, um diskurssemantische Grundfiguren aufzuspüren. Wahrscheinlich wären semantische Annotationen hilfreich, um den Daten eine tiefenstrukturelle Ebene unterlegen zu können. Dagegen lässt sich jedoch mit Recht anmerken, dass damit ein induktiver Zugriff über die Text-



oberfläche aufgegeben und die semantische Annotation Vorannahmen über potenzielle diskurssemantische Grundfiguren erfordern würde.

Die Sprachgebrauchsmuster NICHT NUR ... SONDERN, X IST NICHT MEHR Y, ALS [SELBSTBEZEICHNUNG] HABE ICH und andere mehr, die in der Analyse im Detail untersucht wurden, lassen sich mit den Termini der Argumentationsanalyse genauer beschreiben (vgl. Kapitel 13.2.4, 13.3.2 und 13.3.9). Die Füllungen ihrer Slots stellen (zumindest in argumentativen Kontexten) Topoi dar, deren Vorkommen in Abhängigkeit zu den typischen Sprechweisen in Diskursen steht. Das Beispiel zeigt, wie argumentationsanalytische Kategorien selbst bei einem induktiven, korpuslinguistischen Zugang bei der Einordnung der Phänomene helfen. Aus korpuslinguistischer Sicht ist es dabei besonders interessant, auf textoberflächliche Indikatoren für Topoi zu stoßen. Allerdings müsste sich in einer systematischeren und umfassenderen Analyse erst zeigen, wie viele solcher Indikatoren gefunden werden können. Die berechneten Typikprofile wurden in der vorliegenden Arbeit nur zu einem Bruchteil analysiert und die Analyse fokussierte sich auf die Mehrworteinheiten, die besonders signifikant für jeweils einen Zeitabschnitt sind. Wahrscheinlich sind Indikatoren für Topoi aber eher stabile Sprachgebrauchsmuster und wären in den Typikprofilen in den unteren Signifikanzrängen zu finden, da sie bezüglich ihrer Frequenzen weder für den einen noch den anderen Zeitabschnitt typisch sind, sondern im ganzen untersuchten Zeitraum belegt werden können. Das Typikprofil des Ressorts ‚Leserbriefe‘ (1995–1997 vs. 2003–2005) enthält z. B. die folgenden Mehrworteinheiten, die in beiden Zeiträumen vorkommen und Indikatoren für Topoi sein könnten: *nicht gegen die, ist nicht nur, zum Beispiel der, der des Menschen, die der Welt, Angesichts der der* (alle mit  $G^2 = 0,01$ ), *vor allem die* ( $-0,01$ ), *im Gegensatz zu* ( $-0,05$ ), *scheint zu sein* ( $-0,05$ ), *Tatsache dass die* ( $-0,08$ ).<sup>1</sup> So könnte *Tatsache dass die* als Formel verwendet werden, um Schlüsse vom Faktischen zum Normativen zu ziehen oder *der des Menschen* und *die der Welt* könnten einen GLOBALISIERUNGSTOPOS anzeigen.<sup>2</sup>

Wendet man den sehr weit gefassten Begriff der ‚konzeptionellen Metapher‘ (Lakoff/Johnson 1980) an, lassen sich in den Daten Bei-

<sup>1</sup> Das komplette Typikprofil kann online eingesehen werden: <http://www.bubenhofer.com/korpusanalyse/>.

<sup>2</sup> Vgl. für weitere Sprachgebrauchsmuster in Leserbriefen Kapitel 13.3.9.

spiele finden, die metaphorisch strukturiert sind. So beispielsweise in einiger Komplexität bei den Sprachgebrauchsmustern KAMPF GEGEN X, KAMPF MIT X und KAMPF DEM X. Diese tendieren systematisch zu je spezifischen Füllungen für den Slot X (vgl. Kapitel 13.1.3). Dieser spezifischen Füllungen wegen ergeben sich unterschiedliche Konnotationen der Wendungen *Kampf gegen*, *Kampf mit* und *Kampf dem*. Weil es unüblich ist, den Ausdruck *Kampf dem Terror* zu äußern, aber hingegen üblich *Kampf dem Stau*, birgt das Muster KAMPF DEM X ein spezifisches metaphorisches Potenzial: Die Konnotation der Ausdrucksform *Kampf dem Stau*, z. B. umschrieben als STAU IST EIN LÄSTIGES ÜBEL, überträgt sich auf *Kampf dem Terror* und dieser Ausdruck, *Kampf dem Terror*, wird damit durch die Metapher X IST EIN LÄSTIGES ÜBEL strukturiert. Das Ergebnis ist die Metapher TERROR IST WIE EIN LÄSTIGES ÜBEL. Während im (viel häufigeren) Ausdruck *Kampf gegen (den) Terror* hingegen die Bedeutung eher bei TERROR IST KRIEG liegt, wird in der alternativen Variante ein dem typischen Gebrauch entrissener Ausdruck (*Kampf dem Terror*) durch die Metapher X IST EIN LÄSTIGES ÜBEL strukturiert.

Ein ähnlicher Mechanismus zeigt sich bei der Analyse zu ‚Sorgen und Problemen‘ (vgl. Kapitel 13.2.3 und 14.2). Es sind vornehmlich die Muster KAMPF GEGEN X und BEKÄMPFUNG VON X die beliebige X als Problem strukturieren, das auf eine bestimmte Weise, nämlich ‚kämpferisch‘ gelöst – oder besser: ‚bekriegt‘ werden muss. So wird mit der Aussage *Bekämpfung von Missbräuchen durch illegale Einwanderung oder Kriminalität* (vgl. Beleg 54 auf Seite 254) die ‚illegale Einwanderung‘ mit der Metapher EINWANDERUNG IST WIE EIN ZU BEKÄMPFENDES PROBLEM strukturiert.

In Kapitel 4.6 machte ich neben der konzeptionellen Metapher auf die von Stefanowitsch (2006b) vorgeschlagene ‚Metaphorical Pattern Analysis‘ aufmerksam, die versucht, Metaphern maschinell in Korpora aufzufinden. Ich habe auf eine solche Analyse verzichtet, da sie nicht corpus-driven angewandt werden kann. Es wäre in der corpus-based-Analyse möglich so vorzugehen und ist als Ergänzung zur corpus-driven-Analyse gut vorstellbar.



## 16 Praxis und Theorie

### 16.1 Der Gewinn des korpuslinguistischen Blicks für die Theoriebildung

Die Heuristik, die ich in der vorliegenden Arbeit vorgeschlagen habe (vgl. Teil II), ist in ihrer Vorgehensweise eingeschränkt: Sie geht von der Textoberfläche aus, ohne die sprachlichen Daten vorher zu kategorisieren (z. B. durch eine Lemmatisierung oder eine Auswahl bestimmter Wortarten) und arbeitet blind statistisch. Die überzufällige Rekurrenz von Wörtern ist alleiniges Kriterium, um Mehrworteinheiten aus den Daten zu extrahieren.

Auf der anderen Seite ist die Vorgehensweise so breit wie möglich: Das zu untersuchende Korpus wird in seiner ganzen Breite einbezogen (im NZZ-Korpus werden alle Ressorts berücksichtigt) und es werden alle Mehrworteinheiten extrahiert, die statistisch auffällig sind.

In der zweiten Phase kehrt sich diese corpus-driven-Sicht um und ich versuchte im Rahmen der Heuristik möglichst systematisch auffällige Mehrworteinheiten zu untersuchen, um aus den Daten Sprachgebrauchsmuster zu abstrahieren.

Diese Vorgehensweise ist einerseits theoretisch begründet, führt aber andererseits auch zu neuen Sichtweisen auf ebendiese theoretischen Überlegungen.

#### 16.1.1 Diskurse, Sprechweise, Themen

Diskurse, Sprechweisen und Themen stehen in einem interdependenten Verhältnis. Bestimmte Sprechweisen führen zu bestimmten thematischen Strukturierungen: Wenn ein Thema (z. B. ‚Migration‘, verbalisiert mit unterschiedlichen Lemmata) als Füllung des Slots X in den Mustern KAMPF GEGEN X oder BEKÄMPFUNG VON X verwendet wird, ergibt sich eine spezifische Strukturierung des Themas: MIGRATION IST EIN PROBLEM, MIGRATION MUSS BEKÄMPFT WERDEN etc. So strukturierte Themen führen wiederum zu entsprechend strukturier-

ten diskursiven Praktiken. Im Fall von ‚Migration‘ kann der Diskurs nur noch entlang dieser Struktur geführt werden, die dazu führt, dass Migration entweder als Problem betrachtet wird, oder aber bestritten werden muss, dass es ein Problem ist. Eine dritte Option, z. B. Migration gar nicht erst mit ‚Problem‘ in Verbindung zu bringen, ist nicht mit den Sprechweisen im Diskurs kompatibel.<sup>1</sup>

Für eine Analyse werden nun bestimmte Verhältnisse in den Fokus genommen und die anderen ignoriert. Beispielsweise wird der Zusammenhang zwischen Thema und Diskurs betont, was zu einer Analyse des Diskurses zum Thema X führt. Ich arbeitete vorwiegend zum Verhältnis von Sprechweisen und Diskursen: Ich versuchte ansetzweise zu zeigen, dass Diskurse wie SPRECHEN ÜBER GEWALT oder SPRECHEN ÜBER SORGEN UND PROBLEME beschrieben werden können (vgl. Kapitel 14.1 und 14.2).

Jede Fokussierung auf eines der Verhältnisse führt zu einer Komplexitätsreduktion, die die anderen Aspekte vernachlässigt. Wenn Diskurse über ähnliche Sprechweisen konstruiert werden, gelangt man zu anderen Diskursbeschreibungen, als wenn Diskurse entlang thematischer Kategorien untersucht werden. Im Beispiel der Beschreibung des Migrationsdiskurses oben wird das sichtbar: Die Spezifika des Sprechens über Sorgen und Probleme kann genutzt werden, um den (thematisch definierten) MIGRATIONSDISKURS zu untersuchen. Mit der Fokussierung auf die Sprechweise kann aber genauso gut der Diskurs SPRECHEN ÜBER SORGEN UND PROBLEME untersucht werden, womit die thematische Zentrierung wegfällt und eben die Spezifika dieser Sprechweise analysiert wird. Weil beide Arten von Diskursanalyse Komplexität reduzieren, ist eine Untersuchung aller Verhältnisse zwischen Diskurs, Sprechweise und Themen anzustreben.

### 16.1.2 Kontingente Diskursdefinitionen

Die Definition von Diskursen ist kontingent und ein konstruktiver Akt der Analyse. Daraus ergeben sich drei mögliche Untersuchungsstrategien:

<sup>1</sup> Vgl. dazu auch die Kritik an der Bezeichnung *war on terrorism* im Englischen von Jones/Libicki (2008, 125), erwähnt in Kapitel 14.3.

1. Die Definition eines Diskurses wird als hermeneutischer, subjektiv geprägter Akt angesehen.
2. Es wird versucht, bei der Diskursdefinition Reliabilität zu erreichen, indem vor der Untersuchung Existenzkriterien im Detail festgelegt werden.
3. Es wird versucht, Reliabilität zu erreichen, indem vor der Untersuchung definiert wird, welcher Art die Kristallisationkerne von Diskursen sind. Induktiv werden dieser Definition entsprechend Kristallisationkerne in den Daten sichtbar gemacht. Anschließend werden in einem hermeneutischen Verfahren diese Kristallisationkerne kategorisiert und bilden die Eckpfeiler eines Diskurses.

Dieser ‚dritte Weg‘ verknüpft die beiden anderen Herangehensweisen miteinander. In der vorliegenden Arbeit versuchte ich Letzteres theoretisch zu begründen, methodisch zu operationalisieren und zu exemplifizieren. Als Kristallisationkerne habe ich musterhaften Sprachgebrauch angenommen, der über signifikante Mehrworteinheiten operationalisiert wurde. Diese Operationalisierung ermöglichte die induktive Analyse der Daten und ergab Sprachgebrauchsmuster, die typisch für bestimmte Ressorts, pragmatische Funktionen, Zeitabschnitte oder auch Themen sind.

### 16.1.3 Semantische Matrizen typischer Sprechweisen

Die theoretischen Überlegungen (vgl. Teil I) haben die Verschränkung von Diskursanalyse, Kulturanalyse, Sprachgebrauchsanalyse und weiterer semantisch/pragmatischer Konzepte aufgezeigt. Es stellt sich deshalb die Frage, ob die Sichtweise auf Sprachgebrauch, für die ich plädiere, weiter unter der Bezeichnung ‚Diskursanalyse‘ behandelt werden soll. Die Anschließbarkeit der korpuslinguistischen Diskursanalyse an die anderen theoretischen Konzepte spräche dafür, ‚Diskurs‘ als Spezialfall einer allgemeineren Analysekategorie zu verstehen, nämlich der Kategorie der ‚semantisch geordneten typischen Sprechweisen‘. Diese Kategorie umfasst die folgenden Kriterien:

1. Sie orientiert sich an der sprachlichen Oberfläche.
2. Sie fokussiert Sprachgebrauch, nicht ein ‚Sprachsystem‘.
3. Sie interessiert sich für musterhafte Sprechweisen, die für bestimmte Ausschnitte von Sprachgebrauch typisch sind.
4. Sie geht davon aus, dass diese musterhaften, typischen Sprechweisen semantische Felder strukturieren, also bestimmen, was auf welche Art und Weise wann und wo sagbar und verstehbar ist.

Das Resultat einer von diesen Kriterien geleiteten Untersuchung ist die Beschreibung einer ‚semantischen Matrix‘. Diese ordnet musterhafte, typische Sprechweisen nach beliebigen Kategorien wie ‚Thema‘, ‚kommunikative Gattung‘, ‚Stil‘, ‚Textsorte‘ etc. So könnte eine Kategorie in der semantischen Matrix die Textsorte ‚argumentative Zeitungstexte‘ sein, in der beschrieben ist, welches die musterhaften, typischen Sprechweisen in diesen Daten sind und wie sie die semantischen Felder strukturieren. Weitere Kategorien ergänzten diese Matrix, wobei sich teilweise Überlappungen ergeben würden, wo bestimmte sprachliche Muster für unterschiedliche Kategorien der Matrix typisch sind.

## **16.2 Nutzen einer corpus-driven-Analyse für die Diskurs- und Kulturanalyse**

Mit der Frage nach dem Gewinn des korpuslinguistischen Blicks für die theoretische Konzeption im vorherigen Kapitel habe ich bereits auf einige Konsequenzen für diskurs- und kulturanalytische Fragestellungen hingewiesen. Doch was ergeben sich für Möglichkeiten in der Forschungspraxis?

### **16.2.1 Korpuslinguistik und Diskursanalyse**

Ich plädiere dafür, mit Massendaten zu arbeiten. Das bedeutet keineswegs eine Abkehr von qualitativ ausgerichteten Analyseschritten, wie ich auch in meinen Beispielanalysen gezeigt habe. Eine korpuslinguistische Diskursanalyse muss im Verbund mit anderen diskursanalyti-

schen Verfahren betrieben werden. Allerdings setzt sie an einem grundlegenden Punkt von Diskurslinguistik ein: Eine corpus-driven operierende Korpuslinguistik geht induktiv und damit *hypothese**n**bildend* vor. Statt nur als Hilfsmittel zur Hypothesen**überprüfung** zu dienen, verhilft sie der Diskursanalyse zu einem anderen Startpunkt, indem zunächst ein Korpus auf seinen musterhaften Sprachgebrauch untersucht wird. Die quantitative und qualitative Analyse der Sprachgebrauchsmuster führt alsdann zu bekannten Kategorien wie Metapher, Topos, Argumentationsmuster, Begriff etc., aber auch zu neu zu entwickelnden Kategorien, und ermöglicht die Beschreibung ihrer Funktionen im Diskurs.

Die Sprachgebrauchsmuster bieten darüber hinaus die Möglichkeit, neben einem eher thematischen Fokus auch einen Fokus auf die Sprechweise einzunehmen. Auch hier bietet der induktive Zugang einen Vorteil. Es wäre bedeutend schwieriger deduktiv vorzugehen, da dieser Vorgang eine lexikalisch exakte Definition der Sprachgebrauchsmuster verlangt, die als typisch für einen Diskurs stehen könnten. Der induktive Zugang führt jedoch genau zu den Mustern, die auf einer synchronen oder diachronen Achse für einen bestimmten Sprachgebrauchsausschnitt signifikant sind. Die algorithmische Operationalisierung dieser Muster ermöglicht es, auch umfangreiche Daten auszuwerten.

Aber führt ein Mehr an Daten auch zu einer besseren Diskursanalyse? Ein Mehr an Daten ermöglicht statistisch sicherere Aussagen. Aber der bedeutendere Vorteil eines größeren Datenumfangs zeigt sich in der corpus-driven-Perspektive: Nur damit können jene sprachlichen Veränderungen festgestellt werden, die so gering sind, dass sie unserer Aufmerksamkeit als Leserinnen und Leser entgehen, statistisch jedoch signifikant sind. Diskursive Kräfte können ihre Wirkung auf das Sagbare und auf die Sprechweise auf subtile Art entfalten – die Wirkung ist deshalb nicht weniger stark.

Die quantitative Analyse der Sprachgebrauchsmuster führt aber immer auch zu Resultaten, die keineswegs überraschen. Welche Akteure in der internationalen Politik zu bestimmten Zeiträumen oft in Zeitungen genannt werden, könnte auch ohne quantitative Analyse mit hoher Treffsicherheit vorausgesagt werden. Doch dass die Analysen *auch* zu absehbaren Ergebnissen führen, macht aus der Methode ein



hilfreiches Werkzeug, um diskursanalytische Fragestellungen *empirisch* anzugehen. Damit ist die korpuslinguistische Diskursanalyse anschließbar an alternative sozialwissenschaftliche Methoden, Ergebnisse können verglichen und die Varietät wissenschaftlicher Verfahren erhöht werden.

Die Methode funktioniert unabhängig von der inhaltlichen Gestalt der Daten. In der vorliegenden Untersuchung widerspiegelte das NZZ-Korpus Themen, die für Menschen mit einer durchschnittlichen politischen Allgemeinbildung größtenteils geläufig sind. Werden aber ganz andere Daten, beispielsweise Fachzeitschriften oder Diskussionen in einem subkulturellen Web-Forum analysiert, kann die vorgeschlagene korpuslinguistische Aufbereitung genau so zuverlässig typischen Sprachgebrauch hervorbringen.

Auch wenn eine korpuslinguistische Diskursanalyse induktiv vorgeht, vorurteilslos tut sie es nicht. Wenn der Ausgangspunkt rekurrente lexikalische Elemente sind, die in die Beschreibung von Sprachgebrauchsmustern münden, wird mit diesem Ausgangspunkt bereits eine starke Hypothese als Prämisse gesetzt. Diese Prämisse nimmt solche Sprachgebrauchsmuster als grundlegende Indikatoren für Diskurse an. Die Grenzen einer solchen Diskursanalyse liegen dort, wo eine Gegenhypothese aufgestellt wird. Oder anders ausgedrückt: Auch hier handelt es sich um eine getönte Brille, durch die auf Diskurse geblickt wird. Sie ist m. E. jedoch weniger stark getönt als deduktiv und primär hermeneutisch vorgehende Methoden und bietet damit die Chance, Diskurse zu beschreiben, ohne vorschnell selbst Opfer des Diskurses zu werden.

### 16.2.2 Korpuslinguistik und Kulturanalyse

Während diskurslinguistische Arbeiten oft primär thematisch vorgehen, sind kulturanalytische Arbeiten in der Linguistik meist thematisch unabhängig und eher an Veränderungen im *Sprachgebrauch* interessiert. Denn typischer Sprachgebrauch wird als Mittel gesehen, um die wandelbaren Formen von Kultur zu dechiffrieren. Ich versuchte deshalb mit der vorliegenden Arbeit eine Methode vorzuschlagen, die typischen Sprachgebrauch und dessen Veränderungen in den Blick nimmt. Mit der Möglichkeit induktiv vorzugehen, kann der

thematischen Offenheit von kulturanalytischen Arbeiten entsprochen werden.

Wie ich bereits oben im Kontext der Diskursanalyse betont habe, ersetzt die korpuslinguistische Kulturanalyse keineswegs alternative Verfahren, sondern ergänzt sie um die Möglichkeit, empirisch zu arbeiten und corpus-driven auf typische Muster zu stoßen, die als Hinweise für kulturelle Codes (vgl. Kapitel 2.4.2) oder kommunikative Praktiken im Kontext einer Sprachgebrauchsgeschichte (vgl. Kapitel 4.7) gelesen werden können.

Ähnlich wie bei diskursanalytischen Arbeiten, führt die induktiv vorgehende Korpuslinguistik nicht nur zu unerwarteten, neuartigen Ergebnissen. Doch gilt für kulturanalytische Fragestellungen noch deutlicher, was auch aus diskursanalytischer Perspektive gilt: Es ist nicht zu erwarten, dass sich in elf Jahren Sprachgebrauch in der Neuen Zürcher Zeitung besonders umfassende Veränderungen in kommunikativen Praktiken oder kulturellen Codes widerspiegeln. Dazu ist der Zeitraum zu kurz und die Datenbasis zu wenig vielfältig. Mit der fortschreitenden Digitalisierung wird es bald möglich sein, homogene Textsammlungen quantitativ zu untersuchen, die über mehrere Jahrzehnte reichen. Solche Datenmengen können nur umfassend untersucht werden, wenn quantitative und qualitative Verfahren kombiniert werden.



## 17 Ausblick

### 17.1 Andere Daten, ergänzte Methodik: Ein Beispiel

Zum Schluss möchte ich an einer weiteren kleinen Korpusanalyse skizzieren, wie die vorgeschlagene Methodik weiter verfeinert werden kann und welches Potenzial in der Analyse anderer Daten liegt: Daten aus dem Internetforum von ‚politik.de‘<sup>1</sup>. Aus den Forumstexten stellte ich ein Probekorpus (Politikforum.de-Korpus 2002–2008) zusammen, das mit knapp 49,7 Mio. Wörtern 539 969 Beiträge der 50 aktivsten Autoren umfasst. In diesem deutschsprachigen Forum werden in Unterforen zu allen Bereichen sowohl nationaler als auch internationaler Politik Diskussionen geführt.

Im Unterschied zum NZZ-Korpus habe ich dieses Korpus unter Verwendung des ‚TreeTaggers‘ (Schmid 1994, 1995) mit Wortarten<sup>2</sup> und Lemma-Informationen annotiert. Dann wurden pro Autor die signifikantesten Mehrworteinheiten berechnet.<sup>3</sup>

Schon nach einem oberflächlichen Blick auf die Typikprofile lassen sich autortypische Sprechweisen an den Mehrworteinheiten festmachen. In Tabelle 17.1 auf der nächsten Seite sind für zwei Autoren A und B<sup>4</sup> eine Auswahl der signifikantesten Mehrworteinheiten der beiden Profile aufgeführt. Autor A erinnert mit Wendungen wie *um Prozent auf* (4)<sup>5</sup>, *sagte der Zeitung* (8), *von Milliarden Euro* (9) oder *der gezielten Tötung* (17) an entsprechende Ausdrücke im NZZ-Korpus und damit an Zeitungssprache. Allerdings gibt es auch Ausdrücke

---

<sup>1</sup> Das Forum ist unter <http://forum.politik.de> erreichbar (16. Juli 2008).

<sup>2</sup> Dabei wird das STTS-Tagset verwendet (Schiller u. a. 1995).

<sup>3</sup> Es wurden Dreiworteinheiten auf Basis der Wortformen berechnet; die maximale Distanz zwischen erstem und letztem Token beträgt acht Wörter. Zur Berechnung der Typikprofile wurden dieselben Methoden und Parameter gewählt, wie bereits auf Seite 199 für das NZZ-Korpus beschrieben sind.

<sup>4</sup> Aus Gründen des Personenschutzes verzichte ich auf die Nennung der von den Autoren im Forum verwendeten Selbstbezeichnungen, auch wenn es sich dabei um Spitznamen und nicht reale Namen handelt.

<sup>5</sup> Zahlen in Klammern beziehen sich im Folgenden auf die Zeilennummer der vorher genannten Tabelle.

	G <sup>2</sup>	MWE Autor A		G <sup>2</sup>	MWE Autor B
1	199,65	und Gruss [Spitzname A]	22	893,7	in diesem Land
2	161,25	Bundesagentin für Arbeit	23	574,89	so gut wie
3	134,88	in 100 Jahren	24	400,67	in den Schmerzverstärker
4	126,29	um Prozent auf	25	361,51	Aus diesem Grund
5	110,4	in den nächsten	26	351,23	Ich will dass
6	107,17	in der Atmosphäre	27	333,1	Wichtig ist dass
7	106,42	Prozent als im	28	319,27	meiner Meinung nicht
8	106,29	sagte der Zeitung	29	314,31	Genau das ist
9	95,35	von Milliarden Euro	30	298,96	Sehe ich so
10	92,14	nicht das geringste	31	292,16	zu 100 Prozent
11	92,14	denkt zu viel	32	292,16	ist mal sicher
12	92,14	ist billiger als	33	271,18	ist mir egal
13	82,83	der Erde und	34	258,13	ist meiner Meinung
14	69,11	Den hat man	35	233,72	in meinem Land
15	61,43	Was gespart wird	36	217,03	Mal davon abgesehen
16	61,43	Schon lange hat	37	207,8	Von mir aus
17	61,43	der gezielten Tötung	38	201,66	Man sollte einfach
18	61,43	Was Licht wird	39	193,58	Menschen die nicht
19	61,43	ein wesentlicher die	40	190,67	die ganze Sache
20	61,43	leben rund davon	41	189,76	in jedem Fall
21	61,43	Festhalten an der	42	182,03	die deutsche Sprache
			43	176,54	in der Heimat
			44	175,29	das der Punkt
			45	169,82	in die Heimat
			46	168,89	in keinsten Weise
			47	158,6	heiligt die Mittel
			48	151,81	Wenn es nach
			49	150,84	muss ja nicht
			50	150,25	es mir ginge
			51	141,9	alle Menschen sind
			52	136,26	auf diesem Planeten
			53	125,21	da nur sagen
			54	116,86	Mag schon sein
			55	108,51	den gemeinen Pöbel

**Tabelle 17.1:** Typische Mehrworteinheiten in Beiträgen der Autoren A (links) und B (rechts) im Politikforum.de-Korpus. Manuelle Auswahl für die Analyse (vollständige Version: <http://www.bubenhofer.com/korpusanalyse/>).

wie *nicht das geringste* (10), *denkt zu viel* (11), *Den hat man* (14) etc., die vermuten lassen, dass sie in argumentativen Kontexten verwendet werden.

Bei Autor B sind Ausdrücke, die als Indikatoren für Topoi und Argumentation gelesen werden könnten, noch prominenter als bei Autor A: *Aus diesem Grund* (25), *Ich will dass* (26), *ist mal sicher* (32), *ist meiner Meinung [nicht]* (28, 34), *[der Zweck] heiligt die Mittel* (47) oder *das [ist] der Punkt* (44) sind Beispiele dafür. Zusätzlich werden Lemmata und Ausdrücke genannt, die einen nationalen oder nationalistischen Diskurs vermuten lassen: *in diesem Land* (22), *in meinem Land* (35), *Menschen die nicht* (39), *in der Heimat* (43) usw. Auch Formeln wie *Man sollte einfach [mal]* (38) oder *Wenn es nach [mir ginge]* (48, 50) sind typische Floskeln für politische Laiendiskussionen.

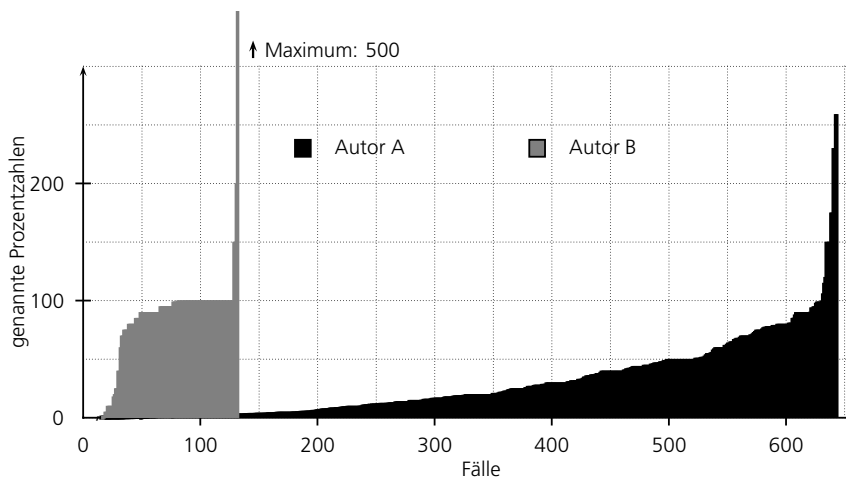
Auf Grund der Verteilung dieser Mehrworteinheiten über das ganze Politikforum.de-Korpus kann rasch festgestellt werden, ob es sich dabei um idiolektale Ausdrücke der Autoren oder eher um soziolektale, thema- oder textsortenspezifische Wendungen handelt. Zusätzlich zeigt sich im Vergleich der beiden Typikprofile mit Typikprofilen des NZZ-Korpus die Nähe von Autor A zur Zeitungssprache. 27,1 % der Mehrworteinheiten von Autor A finden sich auch in den Typikprofilen des NZZ-Korpus (alle Ressorts). Bei Autor B sind es im Vergleich nur 18,8 % aller Mehrworteinheiten.<sup>6</sup>

Wie bereits in den Beispielanalysen zum NZZ-Korpus gibt die Analyse von typischen Slotfüllungen Hinweise auf unterschiedliche Semantiken der Ausdrücke. So verwenden beispielsweise beide Autoren A und B Ausdrücke mit dem Lemma *Prozent* (*um Prozent auf* [4], *Prozent als im* [7], *zu 100 Prozent* [31] etc.). Die Suche nach den spezifischen Füllungen für X im Muster X PROZENT offenbart aber völlig unterschiedliche Verwendungsweisen, wie Tabelle 17.2 auf der nächsten Seite zeigt. Während bei Autor A ganz unterschiedliche Prozentzahlen, auch mit Nachkommastellen vorkommen, scheint die Varianz der genannten Zahlen bei B viel geringer zu sein.

<sup>6</sup> Vgl. die Typikprofile unter [www.bubenhofer.com/korpusanalyse/Politikforum.de/VergleichPolNZZ/](http://www.bubenhofer.com/korpusanalyse/Politikforum.de/VergleichPolNZZ/). Für Autor A gilt: 4882 Mehrworteinheiten kommen sowohl im Typikprofil von A als auch im NZZ-Korpus vor, 13 146 nur im Typikprofil von A. Für Autor B gilt: 2947 Mehrworteinheiten kommen sowohl im Typikprofil von B als auch im NZZ-Korpus vor, 12 743 nur im Typikprofil von B.

Autor A		
1	Praxis der Rückgabe aber für zu kompliziert.	79 Prozent
2	gegeben werden können, wo sie gekauft wurden.	65 Prozent
3	itschrift «Auf einen Blick». Demnach kaufen	60 Prozent
4	ckungen. Unter den 14- bis 29-Jährigen lehnen	48 Prozent
5	dinap zufolge kommen die Christsozialen auf	62 Prozent
6	um rechnet. Im nächsten Jahr hält der Fonds	1,5 Prozent
7	Prozent Wachstum für machbar, neben Japan (	1,4 Prozent
8	pt. IWF erwartet für 2003 und 2004 jeweils	3,9 Prozent
9	t einem gesamtstaatlichen Defizit von jeweils	3,9 Prozent
10	vor kurzem eine Defizitprognose für 2003 von	3,8 Prozent
11	Praxis der Rückgabe aber für zu kompliziert.	79 Prozent
12	gegeben werden können, wo sie gekauft wurden.	65 Prozent
13	aget erleidet schwere Verluste und sackt auf	18,7 Prozent
14	islang teilnehmenden Staaten stellen lediglich	44 Prozent
Autor B		
15	ste! Was für eine grandiose Aussage! Auch	85 Prozent
16	sere Rechtsprechung falsch hoch zehn ist. Bei	80 Prozent
17	che Körperverletzung. Als Asylant hat man zu	100 Prozent
18	gar dann noch widersprechen, wenn ich ihm zu	100 Prozent
19	straffällig wurde! Ich denke mal, daß etwa	95 Prozent
20	t erst dann für die EU akzeptabel, wenn nicht	10 Prozent
21	! Für mich kann man Stadtteile mit mehr als	20 Prozent
22	lländischen oder französischen Einwanderern zu	99 Prozent
23	ebären. Du hast die Wahrheit denke ich zu	99,9 Prozent
24	auch zusätzlich mit der FDP) wie in Bayern.	18 Prozent
25	Gutmenschen und Rassisten, sondern auch etwa	95 Prozent
26	ern, die pro Jahr nach Deutschland einwandern	80 Prozent
27	, die aus der Türkei oder aus Albanien kommen	80 Prozent
28	en erregende Quote von Menschen liegt bei etwa	75 Prozent

**Tabelle 17.2:** Auswahl von Belegen für das Muster X PROZENT in den Beiträgen der Autoren A und B im Politikforum.de-Korpus.



**Abbildung 17.1:** Alle genannten Prozentzahlen im Muster X PROZENT in den Beiträgen der Autoren A und B im Politikforum.de-Korpus. Lesehilfe: Autor B nannte etwa 50 Mal die Prozentzahl 100, bei Autor B kommt 100 hingegen kaum vor.

Der Blick in die Belege macht deutlich, dass die Verwendung des Musters X PROZENT bei den beiden Autoren völlig unterschiedlich ist. Während bei Autor A der *exakte* Wert der Prozentzahl der Argumentation dient, sind bei Autor B die Werte höchstens als Schätzungen zu verstehen oder werden verwendet, um Überzeugung oder Sicherheit auszudrücken (*Du hast die Wahrheit denke ich zu 99.9 Prozent erkannt*). Diese unterschiedlichen Verwendungsweisen sind bei den beiden Autoren systematisch, wie Abbildung 17.1 zeigt. Dort ist aufgeführt, welche Prozentzahlen von den Autoren wie oft verwendet wurden. Während bei A eine kontinuierlich ansteigende Kurve sichtbar ist, also ganz unterschiedliche Zahlen zwischen 0 und 100 (sowie wenigen Zahlen über 100) genannt werden, verläuft die Kurve bei B abrupter und gestufter.

Diese Befunde über die unterschiedlichen Sprechweisen der beiden Autoren lassen sich nun ergänzen durch weitere Analysen. Da das Politikforum.de-Korpus mit Wortarten annotiert und lemmatisiert ist, lassen sich beispielsweise die je Autor typischen Lemmata ausgeben, beschränkt auch auf bestimmte Wortarten. In Tabelle 17.3 auf der nächsten Seite sind für die beiden Autoren die je 30 signifikantesten



Autor A (Rang nach G <sup>2</sup> )				Autor B (Rang nach G <sup>2</sup> )			
<b>Nomen</b>				<b>Nomen</b>			
1	Jahr	16	Mrd.	1	Mensch	16	Caution
2	%	17	Bahn	2	Kommunist	17	Urknall
3	Energie	18	Meter	3	Ausländer	18	Redaktion
4	Strom	19	Öl	4	Türke	19	Kopftuch
5	Prozent	20	Windenergie	5	Religion	20	Frau
6	Milliarde	21	Kosten	6	Deutsche	21	Penner
7	Euro	22	Angabe	7	Islam	22	Einwanderer
8	CO <sub>2</sub>	23	kWh	8	D	23	Kultur
9	Tonne	24	Fläche	9	Forum	24	Leben
10	Subvention	25	Windkraftan- lage	10	Typ/Typus	25	Schmerz- verst...
11	Anlage	26	Cent	11	Kind	26	Moslem
12	Kraftwerk	27	Preis	12	Land	27	Gesellschaft
13	Windrad	28	Technologie	13	Beitrag	28	Täter
14	Kernkraft- werk	29	Steuer	14	Kommunis- mus	29	Sprache
15	Million	30	Wasser	15	Meinung	30	Gott
<b>Verben</b>				<b>Verben</b>			
1	berichten	16	fahren	1	schreiben	16	gehören
2	steigen	17	verbrauchen	2	sprechen	17	sehen
3	bauen	18	ergeben	3	denken	18	melden
4	kosten	19	mitteilen	4	sagen	19	gebären
5	rechnen	20	zahlen	5	glauben	20	tun
6	erklären	21	betonen	6	finden	21	leben
7	sagen	22	nutzen	7	sperrern	22	meinen
8	sinken	23	decken	8	bestrafen	23	abschieben
9	erhöhen	24	wachsen	9	verstehen	24	wissen
10	betragen	25	zeigen	10	leiden	25	fühlen
11	erzeugen	26	finden	11	lernen	26	gefallen
12	liefern	27	fördern	12	kennen	27	arbeiten
13	entdecken	28	erhöhen	13	scheißen	28	reden
14	subventionie- ren	29	untersuchen	14	integrieren	29	benehmen
15	liegen	30	sparen	15	hassen	30	machen
<b>Adjektive</b>				<b>Adjektive</b>			
1	teuer	16	hoch	1	original	16	wichtig
2	billig	17	technisch	2	einfach	17	schwul
3	neu	18	künftig	3	deutsch	18	genau
4	vergangen	19	warm	4	türkisch	19	recht
5	garnicht	20	international	5	undeutsch	20	nichtmal
6	knapp	21	wirtschaftlich	6	dunkel	21	asozial
7	russisch	22	europäisch	7	wirklich	22	primär
8	jährlich	23	deutlich	8	gut	23	cool
9	fossil	24	erneuerbaren	9	moralisch	24	ausländisch
10	radioaktiv	25	kanadisch	10	schlimm	25	dumm
11	tatsächlich	26	britisch	11	kriminell	26	gewiß
12	geplant	27	steigend	12	mies	27	egal
13	weltweit	28	chinesisch	13	echt	28	muslimisch
14	erst	29	gering	14	religiös	29	stolz
15	gigantisch	30	regenerativ	15	islamisch	30	christlich

**Tabelle 17.3:** Die je 30 typischsten Nomen, Verben und Adjektive in Beiträgen der Autoren A (links) und B (rechts) im Politikforum.de-Korpus (Ränge nach abnehmender Signifikanz; vollständige Version: <http://www.bubenhofer.com/korpusanalyse/>).

Nomen, Verben und Adjektive aufgeführt. Einerseits lassen sich sofort thematische Prioritäten in den Beiträgen der Autoren erkennen. Bei A geht es um Energiepolitik, bei B um Migration. Auf Grund der je spezifischen Verben und Adjektive lassen sich rasch Unterschiede im Diskussionsstil ableiten. So gibt es unter den typischsten Adjektiven von B viele Expressiva wie *schlimm*, *mies*, *cool*, *dumm* oder *stolz*, während die Adjektive von A neutralere Konnotationen aufweisen. Die Daten zeigen, dass es sich lohnt, weitere quantitative Analyseverfahren auf die Daten anzuwenden, z. B. indem gezielt nach Intensivierungen wie Grad- und Fokuspartikeln (*absolut*, *gänzlich*, *komplett*, *höchst*, *mega*, *derart*, *ziemlich*, *etwas*, *kaum*, *beinahe*, *nicht im geringsten* etc.) gesucht wird.

Ich verzichte an dieser Stelle auf eine detaillierte Analyse und möchte stattdessen nochmals auf das Potenzial eingehen, dass sich bei diesen Daten zeigt:

- Mit den in der vorliegenden Arbeit entwickelten Methoden lassen sich auch in einem Web-Diskussionsforum rasch vorherrschende Diskurse aber auch typische Sprechweisen aufdecken. Während das NZZ-Korpus einen stark kodifizierten Sprachgebrauch widerspiegelt, sind die Variationen im Sprachgebrauch im Politikforum.de-Korpus breiter und das Korpus hinsichtlich bestimmter Fragestellungen, wie z. B. Formen der Argumentation, ergiebiger.
- Die Annotation des Korpus mit Wortarten und Lemmata ermöglicht ergänzende Analysen der Daten. Für die Berechnung von Mehrworteinheiten (vgl. Tabelle 17.1 auf Seite 326) ist eine Annotation nicht zwingend. Um jedoch Schlüsselwortanalysen zu betreiben oder bestimmte syntaktische Analysen vorzunehmen, ist eine Annotation sehr hilfreich.

Mittels einer ganzen Reihe von Oberflächenphänomenen wie Mehrworteinheiten, Sprachgebrauchsmustern, Schlüsselwörtern, Expressiva, Intensivierungsindikatoren etc. könnten für Teilkorpora (Autoren, Autorengruppen, Textgruppen usw.) Sprachgebrauchsprofile erstellt werden, die sich durch je spezifische Korrelationen dieser Phänomene auszeichnen.

## 17.2 Schlüsse, Möglichkeiten, Desiderata

Aus dem Beispiel im vorherigen Kapitel lassen sich zusammenfassend einige Schlüsse für weitere Untersuchungen ziehen und es werden Ergänzungen der Methode, wie ich sie in der vorliegenden Arbeit entwickelt habe, deutlich. Zudem möchte ich auf einige methodische Desiderata hinweisen.

### 17.2.1 Korpora

Eine diachrone Perspektive auf den Publikationszeitraum einer Tageszeitung von 1995 bis 2005 lässt keine starken Umwälzungen im Sprachgebrauch vermuten. Erschwerend kommt hinzu, dass die Neue Zürcher Zeitung ein eher konservatives Image pflegt und in diesem Zeitraum wenige bewusste Änderungen im Stil, Layout und in der Struktur vornahm. Es handelt sich also um einen stark kodifizierten und reglementierten Sprachgebrauch.

Die quantitativen Analysen haben trotzdem Veränderungen im Sprachgebrauch zu Tage gebracht. Dies unterstreicht das heuristische Potenzial empirischer Analysen. Interessant wird aber auch die Analyse von alternativen Korpora sein. Die Medienwirklichkeit ist eine unter vielen Wirklichkeiten; um eine umfassendere Beschreibung von Diskursen oder kulturellen Veränderungen zu erreichen, ist die Analyse anderer Textdaten unumgänglich.

Beinahe jede Textquelle, die eine gewisse serielle Homogenität aufweist, ist für ähnliche Untersuchungen interessant. Oft sind es forschungspraktische Probleme, die eine quantitative Analyse erschweren: Die Texte müssen digitalisiert und kategorisiert, Urheberrechte geklärt werden. Es gibt aber eine Menge an Textdaten, die vergleichsweise einfach genutzt werden kann.<sup>7</sup> Neben Tageszeitungen denke ich an die folgenden Datenquellen:

- Bereich Massenmedien:
  - Zeitschriften aus den Bereichen Mode, Freizeit, Hobby, Politik etc.
  - Nicht-etablierte Medien wie Blogs, Podcasts etc.

---

<sup>7</sup> Vgl. z. B. für politische Textsorten die Analysen zu den US-Wahlen 2008 und den deutschen Bundestagswahlen 2009: Bubenhofer u. a. (2008a,b, 2009).

- Bereich interaktive Kommunikationsplattformen:
  - Web-Foren, Mailing-Listen, Plauderplattformen etc.<sup>8</sup>
  - Webplattformen zur persönlichen Präsentation und Bildung sozialer Netzwerke
  - Leser/innen-Kommentare bei Online-Medien, Blogs etc.
- Institutionen- und Unternehmenskommunikation:
  - Öffentliche Protokolle von Parlamenten und politischen Gremien<sup>9</sup>
  - PR-Erzeugnisse von Unternehmen, Parteien, Organisationen etc.
  - Unternehmensinterne Kommunikation in Form von E-Mails, Protokollen etc.

Zudem schreitet die Digitalisierung älterer Texten rasch voran, so dass sich die Situation für korpuslinguistische Diskursanalysen kontinuierlich verbessert.

Wichtig ist jedoch die exakte Situierung der verwendeten Quellen im ‚allgemeinen Sprachgebrauch‘. Es muss genau definiert werden, welchen Sprachgebrauchsausschnitt das Korpus repräsentiert und wie groß die entsprechende Grundgesamtheit ist, um valide Aussagen machen zu können (vgl. dazu auch Kapitel 5.3).

### 17.2.2 Statistik

Die in der vorliegenden Arbeit benutzten statistischen Methoden sind einfach. Dies hat den Vorteil, dass sie auch von nicht statistisch ausgebildeten Personen nachvollzogen und angewandt werden können. Die Gemeinde der empirisch ausgerichteten Korpuslinguistinnen und -linguisten arbeitet zudem an der Verfeinerung des statistischen Methodenapparates und dessen Anpassung an korpuslinguistische Bedürfnisse. Da stellen sich grundlegende Fragen zu statistischen Prämissen, die für sprachliche Daten nur bedingt gelten. Erwähnt sei an

---

<sup>8</sup> In Kapitel 17.1 der vorliegenden Arbeit habe ich Analysemöglichkeiten in einem Diskussionsforum zu politischen Themen skizziert.

<sup>9</sup> In der vorliegenden Arbeit verwendete ich z. B. die elektronisch verfügbaren Wortprotokolle des eidgenössischen Parlaments (Amtliches Bulletin o. J.).

dieser Stelle das Problem der Klumpenhaftigkeit von sprachlichen Daten (vgl. Kapitel 7.2.1).

Statistische Methoden sind in der diskurs- und kulturanalytischen Linguistik von großem Nutzen. Deshalb ist es nötig, dass die quantitativen Desiderata formuliert werden, damit statistische Standardverfahren definiert werden können. Diese Verfahren müssen einfach auf Korpusdaten anwendbar sein, ohne dass der statistische Hintergrund im Detail gekannt werden muss. Im Idealfall fließen solche Verfahren in korpuslinguistische Computerprogramme ein, die unkompliziert bedient werden können.

Es gibt eine Reihe geeigneter Software für solche Zwecke, die jedoch häufig schwierig zu bedienen ist, nicht mit sehr großen Datenmengen umgehen kann oder nicht die gewünschten Funktionen aufweist. Aus eigener Erfahrung wünsche ich mir ein Korpusanalyseprogramm, das im Groben folgende Funktionen aufweist:

- Verwaltung von umfangreichen Korpora in einer Datenbank, die auch alle verfügbaren Metainformationen enthält.
- Importfilter, um bestehende Daten einlesen und nach Metainformationen absuchen zu können. Leider liegen die gewünschten Daten oft nicht in einem strukturierten Format wie XML vor, sondern im Rohtext.
- Eingebaute Lemmatisierungs- und Annotationswerkzeuge.
- Werkzeug zur corpus-driven Clusterberechnung aufgrund verschiedener Kriterien.
- Statistische Angaben zur Distribution von beliebigen Phänomenen im Korpus.
- Umfangreiche Recherche- und Trefferdarstellungsmöglichkeiten.

Es bleibt aber anzumerken, dass auch mit simplen Methoden Analysen möglich sind und mit der existierenden Software nach einer gewissen Einarbeitungszeit nützliche Analysewerkzeuge vorliegen (vgl. auch Kapitel 9).

### 17.2.3 Semantische Kategorisierungen und Annotation

In Kapitel 6.5 habe ich mich kritisch gegenüber der Annotation von Korpusdaten mit Lemmata und Wortarten ausgesprochen. Die Analysen (vgl. Teil III) haben dieser kritischen Einstellung grundsätzlich nicht widersprochen. Sie haben aber auch gezeigt, dass Kategorisierungen nach semantischen Kriterien, aber auch Lemmatisierungen und Wortartenannotationen punktuell nützlich wären, um in der corpus-based-Analyse Hypothesen zu überprüfen.<sup>10</sup> Aus diesen Erfahrungen ergeben sich zwei Schlussfolgerungen:

1. Die Annotation von Lemmata und Wortarten, aber auch der Syntax, könnte die corpus-based-Analyse unterstützen. Nicht sinnvoll ist sie bei der corpus-driven-Berechnung von typischen Mehrworteinheiten. Aber eine geschickte Annotation ermöglicht es problemlos, die Annotation nach Belieben zu nutzen oder zu ignorieren.
2. Es gäbe weitere Kategorien, die für eine Annotation interessant sein könnten. Eine semantische Kategorisierung von Ausdrücken, die z. B. positive oder negative Konnotationen erfasst, könnte die Analysemöglichkeiten um wichtige Aspekte erweitern. Oder es könnten Erkenntnisse über Indikatoren für Argumentationsmuster auf der Textoberfläche genutzt werden, um im Korpus Slots in solchen argumentativen Strukturen zu markieren.

Teilweise existieren bereits computerlinguistische Konzepte und Methoden, um solche semantischen Annotationen vorzunehmen. Auch lexikalisch-semantische Wortnetze wie ‚GermaNet‘ sind mögliche Ressourcen für solche Vorhaben.<sup>11</sup> Allerdings ist es wichtig, dass von linguistischer Seite deutlich artikuliert wird, welche Form von Annotationen und Kategorisierungen gewünscht sind. Denn die Computerlinguistik versteht sich zu Recht nicht als ‚Dienstleister‘ für andere Wissenschaften, sondern verfolgt eigene Erkenntnisinteressen, die sich aber nur bedingt mit den Forschungsinteressen einer diskurs- und

<sup>10</sup> Vgl. dazu auch Kapitel 17.1.

<sup>11</sup> Vgl. für einen Überblick Carstensen u. a. (2001, 246ff.) und speziell zu ‚GermaNet‘: Carstensen u. a. (2001, 387).

kulturanalytisch ausgerichteten Linguistik decken. Eine gemeinsame Beschäftigung mit Themen und Methoden wäre mit Sicherheit für beide Seiten bereichernd.<sup>12</sup>

#### 17.2.4 Sprachgebrauchsmuster als kleinste semantische Einheiten

Die Phraseologie, gebrauchorientierte Semantiktheorien und die Pragmatik im Allgemeinen haben schon früh den Blick über die Wortgrenzen hinweg auf Phrasen gelenkt, um deren Semantik und pragmatischen Funktionen zu beschreiben (vgl. Kapitel 2.1, 3.1.1 und 4.4). Diese Sichtweise hat m. E. in der Diskursanalyse noch zu wenig Fuß gefasst; deshalb war es auch eines der Ziele der vorliegenden Arbeit, den Fokus auf Sprachgebrauchsmuster zu lenken, die die Wortgrenze sprengen.

Einigt man sich grundsätzlich darauf, Mehrworteinheiten als Basis für semantische Analysen im weiteren Sinn zu verwenden, bleibt noch immer offen, wie diese Mehrworteinheiten definiert sein sollen. In der phraseologischen Forschung konzentrierte man sich auf Sprichwörter, Redewendungen oder Phraseme, die sich durch Idiomatizität auszeichnen (Burger 1998, 50f.). Auch in der Korpuslinguistik gilt die Aufmerksamkeit häufig primär den ‚Kollokationen‘, die idiomatischen Charakter haben (vgl. Kapitel 6). Die Tauglichkeit statistischer Maße werden demnach an solchen Definitionen gemessen.<sup>13</sup> Ich möchte jedoch im Anschluss an Feilke (1993) dafür plädieren, für semantische Analysen keinen solch eingeschränkten Kollokationenbegriff zu verwenden. Vielmehr ist von Mehrworteinheiten auszugehen, die grundsätzlich jede Kombination von Wörtern umfasst, die typisch für bestimmte Sprachausschnitte ist. Nur so ist es möglich, Sprachgebrauch und dessen Wandel im Sinne von Linke (2003a) in seiner ganzen Fülle zu erfassen. Somit werden auch Routineformeln, Funktionsverbgefüge, schlicht irgendwelche Satzfragmente erfasst, die musterhaft sind.

<sup>12</sup> Die gezeigten Kontextualisierungsleistungen von Sprachgebrauchsmustern könnten Anregungen zur Lösung computerlinguistischer Probleme bieten: So wäre es vorstellbar, für bestimmte Textsorten typische Sprachgebrauchsmuster zur automatischen Textsortenbestimmung zu nutzen oder semantische Annotationsverfahren könnten um pragmatische Aspekte ergänzt werden. Vgl. dazu Bubenhofer/Scharloth (2009).

<sup>13</sup> Vgl. z. B. Pearce (2002) und Fußnote 18 auf Seite 123.

Auf diese Weise können Sprachgebrauchsmuster als kleinste semantische Einheiten nicht nur die Basis für diskurs- und kulturanalytische Fragestellungen bilden, sondern sind zugleich für semantische und pragmatische Perspektiven auf Sprachgebrauch nützlich. Dass Sprachgebrauchsmuster für empirische Untersuchungen operationalisierbar sind, hoffe ich in meiner Untersuchung gezeigt zu haben.





V

Anhang



## A Typikprofile

### A.1 Ressort ‚Ausland‘

#### A.1.1 Periode 1995–1997

Zeile	G <sup>2</sup>	MWE	Zeile	G <sup>2</sup>	MWE
1	42.15	Human Rights Watch	46	19.76	die Fortsetzung der
2	39.52	und der Opposition	47	16.07	die Sicherheit der
3	15.81	Russland und der	48	15.81	die Freilassung der
4	32.93	der der Menschenrechte	49	22.39	in New Hampshire
5	19.76	der der Bundeswehr	50	22.39	sich die Frage
6	17.12	der der Slowakei	51	21.07	den Reihen der
7	15.81	der der Herzegowina	52	15.81	den der letzten
8	15.81	der der republikani-	53	15.81	den Status der
		schen			
9	22.87	der Hauptstadt der	54	21.07	in den Reihen
10	22.39	der fruehere der	55	21.07	und in Moskau
11	22.29	der Partei der	56	15.81	und noch in
12	21.07	der Konferenz der	57	19.76	der sowie des
13	31.61	haben am Donnerstag	58	19.76	der anderen Seite
14	30.3	San Suu Kyi	59	19.76	Karadzic und Mladic
15	35.56	Aung San Suu	60	18.44	zwischen und des
16	30.3	der ersten Runde	61	18.44	Hinblick auf die
17	27.66	vor den Wahlen	62	18.44	bosnischen Serben die
18	27.66	die immer wieder	63	47.15	der bosnischen Ser-
					ben
19	27.66	gegen der Regierung	64	27.9	die bosnischen Ser-
					ben
20	26.34	in der Slowakei	65	18.44	erster Linie der
21	22.39	in der Politik	66	18.44	der Republik die
22	22.39	in der Oeffentlich-	67	17.12	der die politische
		keit			
23	21.07	in der Foederation	68	18.44	mehr oder weniger
24	17.12	in der Umgebung	69	18.44	mit der Tuerkei
25	17.12	in der Kammer	70	18.44	die Wahlen in
26	15.81	in der DDR	71	18.44	die in Mostar
27	21.07	in Linie der	72	17.12	die in Regierung
28	19.76	in Sarajewo der	73	17.12	vor einer Woche
29	15.81	in Genf der	74	17.12	Kim Dae Jung
30	17.12	letzten in der	75	17.12	die Wege geleitet
31	15.81	Zeit in der	76	17.12	in Sarajewo und
32	26.34	am abend der	77	15.81	in Russland und
33	21.07	am Donnerstag abend	78	17.12	auf der Strasse

**Tabelle A.1:** (Fortsetzung folgende Seite)

**Tabelle A.1:** (Fortsetzung von vorheriger Seite)

Zeile	G <sup>2</sup>	MWE	Zeile	G <sup>2</sup>	MWE
34	25.36	der Nato und	79	15.81	gegen die Korruption
35	15.81	der Nato zu	80	15.81	der Zone Nord
36	22.34	die der Nato	81	15.81	im Parlament der
37	25.03	die die Nato	82	15.81	in Bosnien die
38	18.44	die Nato der	83	31.71	der in Bosnien
39	23.71	in die Nato	84	15.81	Kim Young Sam
40	25.03	am in Paris	85	15.81	Sao Felix Araguaia
41	23.71	die absolute Mehrheit	86	15.81	Bundeskanzler Kohl in
42	23.71	russische Praesident Jelzin	87	15.81	Praesident Clinton und
43	22.39	die Koalition der	88	15.65	die Tatsache dass
44	18.44	die der Wahlen	89	15.01	im Rahmen des
45	15.81	die der Bundeswehr			

**Tabelle A.1:** Typische Mehrworteinheiten (Typikprofil) im Ressort ‚Ausland‘ des NZZ-Korpus in den Jahren 1995–1997 im Vergleich zu den Jahren 2003–2005 (manuell gefiltert, Cluster,  $G^2 \geq 15$ ).

## A.1.2 Periode 2003–2005

Zeile	G <sup>2</sup>	MWE	Zeile	G <sup>2</sup>	MWE
1	−72.89	im Irak die	57	−26.24	den besetzten Gebieten
2	−48.11	im Irak und	58	−26.24	die Opposition der
3	−48.11	im Irak der	59	−17.49	die Bevoelkerung
					der
4	−39.36	im Irak zu	60	−16.04	die Opfer der
5	−32.07	im Irak in	61	−24.78	man sich in
6	−20.41	im Irak den	62	−24.78	der Internationalen
					Atomenergieagentur
7	−16.04	im Irak von	63	−24.78	im Weissen Haus
8	−164.74	der im Irak	64	−24.78	die Kontrolle ueber
9	−153.08	die im Irak	65	−24.78	auf in Bagdad
10	−61.23	den im Irak	66	−23.33	der Ministerpraesident
					Sharon
11	−59.77	und im Irak	67	−23.33	in neunziger Jahren
12	−27.7	amerikanischen	68	−16.48	in vergangenen
		im Irak			Jahren
13	−27.7	fuer im Irak	69	−21.87	Usama bin Ladin
14	−21.87	auf im Irak	70	−21.87	in letzten von
15	−43.74	aus dem Irak	71	−24.78	und in letzten
16	−33.53	aus dem Gazast-	72	−21.87	den der Hauptstadt
		reifen			
17	−48.11	Abzug aus dem	73	−20.41	Sánchez de Lozada
18	−43.14	den USA und	74	−20.41	im Sueden des
19	−30.62	den USA die	75	−20.41	in immer wieder
20	−17.49	den USA zu	76	−20.41	Praesident und die
21	−52.48	der den USA	77	−20.3	der Suche nach
22	−26.63	die den USA	78	−18.95	sagte ein der
23	−21.31	mit den USA	79	−18.95	in besetzten Gebieten
24	−41.82	gegen den Terrorismus	80	−18.95	vor allem mit
25	−40.82	gegen den Irak	81	−18.95	Debatte ueber die
26	−36.45	den gegen den	82	−18.95	das zwischen den
27	−24.49	Kampf gegen den	83	−18.95	Der amerikanische
					Praesident
28	−21.87	Krieg gegen den	84	−18.95	der und Uno
29	−21.87	Kampf gegen	85	−16.04	der und Gerech-
		Terrorismus			tigkeit
30	−17.49	im gegen Terroris-	86	−18.95	Kampf den Terroris-
		mus			mus
31	−37.91	den neunziger Jahren	87	−18.2	der neunziger Jahre
32	−35.28	hiess es in	88	−17.49	die Amerikaner die
33	−29.16	Praesident Bush in	89	−17.49	Irak Neue Zuercher
34	−21.87	Praesident Bush	90	−17.49	einem der Stadt
		hat			
35	−17.49	Praesident Bush	91	−17.49	im den Terrorismus
		den			
36	−17.49	Praesident Bush	92	−17.49	von Johannes Paul
		die			
37	−17.49	von Praesident	93	−16.04	von Johannes II
		Bush			
38	−29.16	der fuer einen	94	−16.04	von Paul II

Tabelle A.2: (Fortsetzung folgende Seite)

**Tabelle A.2:** (Fortsetzung von vorheriger Seite)

Zeile	G <sup>2</sup>	MWE	Zeile	G <sup>2</sup>	MWE
39	-29.16	die USA in	95	-16.48	eine Loesung des
40	-18.95	die in Darfur	96	-16.04	der gemaessigten
41	-18.95	die amerikani-	97	-16.04	Islamisten
42	-36.45	schen in	98	-16.04	seine Partei der
43	-25.06	dass die USA	99	-16.04	Zahl der auf
44	-17.49	und die USA	100	-17.49	Saddam Hussein der
45	-29.16	fuer die USA	101	-16.04	die Saddam Huss-
46	-26.32	in der irakischen	102	-16.04	eins
47	-18.95	in der amerikani-	103	-16.04	von Saddam
48	-16.04	schen	104	-30.62	Hussein
49	-29.16	Sharon in der	105	-16.04	israelische Minister-
50	-25.06	Bush in der	106	-16.04	praesident Sharon
51	-16.15	der der politischen	107	-18.95	in Kosovo und
52	-21.87	der der amerikani-	108	-17.49	der in Kosovo
53	-27.7	schen	109	-16.04	der Partei Gerechtig-
54	-27.7	Kritik der der	110	-15.3	keit
55	-68.52	Côte d Ivoire			die EU den
		den Irak zu			der die EU
		in den Irak			in die EU
					an die Oeffentlichkeit
					hat den letzten

**Tabelle A.2:** Typische Mehrworteinheiten (Typikprofil) im Ressort ‚Ausland‘ des NZZ-Korpus in den Jahren 2003–2005 im Vergleich zu den Jahren 1995–1997 (manuell gefiltert, Cluster,  $G^2 \geq 15$ ).

## A.2 Ressort ‚Inland‘

## A.2.1 Periode 1995–1997

Zeile	G <sup>2</sup>	MWE	Zeile	G <sup>2</sup>	MWE
1	962.54	Ja Ja Ja	65	26.22	Angesichts der der
2	224.71	Ja Ja Nein	66	21.22	Zusammenarbeit der
3	49.94	Ja Ja Partei	67	18.73	der
4	46.19	Ja Ja Freigabe	68	16.23	will der der
5	34.95	Ja Ja Freiheitspartei	69	36.20	Beruecksichtigung
6	34.95	Ja Ja ACS	70	21.22	der der
7	26.22	Ja Ja SPS	71	62.42	Nein Nein Ja
8	26.22	Ja Ja Gruene	72	32.46	Nein Nein Nein
9	26.22	Ja Ja EVP	73	19.97	Ja Nein Nein
10	26.22	Ja Ja SVP	74	18.73	Einheit der Materie
11	18.73	Ja Ja KVP	75	32.46	der Einheit Materie
12	18.73	Ja Ja TCS	76	32.46	die Einheit Materie
13	18.73	Ja Ja LdU	77	32.46	der Lex Friedrich
14	18.73	Ja Ja Liberale	78	31.21	aus der in
15	148.56	Ja Nein Ja	79	15.49	Jugend ohne Drogen
16	54.93	Ja Partei Ja	80	19.85	Verhandlungen mit EU
17	34.95	Ja EVP Ja	81	19.97	Verhandlungen mit
18	31.21	Ja SVP Ja	82	29.96	Verhandlungen der
19	29.96	Ja SPS Ja	83	27.46	EU
20	28.71	Ja LdU Ja	84	24.97	die Verhandlungen
21	27.46	Ja Gewerkschafts- bund Ja	85	27.46	mit
22	27.46	Ja CVP Ja	86	17.48	der Telecom PTT
23	26.22	Ja Freigabe Ja	87	26.22	mit gegen Stimmen
24	22.47	Ja Liberale Ja	88	33.71	wird mit Stimmen
25	22.47	Ja Gruene Ja	89	24.97	des Zweiten Weltkriegs
26	18.73	Ja Freiheitspartei Ja	90	23.72	Schatten des Zweiten
27	16.23	Ja Vorort Ja	91	23.72	Schweiz Zweiten Welt-
28	151.06	Nein Ja Ja	92	23.72	krieg
29	56.18	Partei Ja Ja	93	22.47	der Schweiz Zweiten
30	34.95	Vorort Ja Ja	94	21.22	Bundesraetin Ruth Drei-
31	34.95	LdU Ja Ja	95	19.97	fuss
32	34.95	SVP Ja Ja	96	15.15	fuer ein der
33	31.21	Freigabe Ja Ja	97	19.97	befasst sich der
34	26.22	FDP Ja Ja	98	18.73	von Otto Stich
35	26.22	CVP Ja Ja	99	18.73	die des neuen
36	26.22	Parteien Ja Ja	100	18.73	die zur Verfuegung
37	18.73	SPS Ja Ja	101	18.73	der Bevoelkerung die
38	18.73	EVP Ja Ja	102	18.73	der gegen die

Tabelle A.3: (Fortsetzung folgende Seite)



**Tabelle A.3:** (Fortsetzung von vorheriger Seite)

Zeile	G <sup>2</sup>	MWE	Zeile	G <sup>2</sup>	MWE
39	18.73	Gruene Ja Ja	103	18.08	mit der werden
40	18.73	Bahnhöfen Ja Ja	104	20.48	Zusammenhang mit
41	18.73	Gewerbeverband Ja	105	17.48	der
		Ja			im Sinne des
42	18.73	Liberale Ja Ja	106	17.48	Grund der und
43	42.45	bezug auf die	107	117.35	auf Grund der
44	57.43	in bezug auf	108	18.73	Auf Grund der
45	41.20	Eidgenössische Räte	109	17.48	die Regierung und
		rom			
46	38.70	Eidgenössische	110	17.48	die Zusammenarbeit
		Räte M			und
47	38.70	im Zweiten Weltkrieg	111	17.48	in der Praxis
48	36.20	der im Zweiten	112	17.48	Kurt Schuele Schaffhau-
					sen
49	26.22	Schweiz im Zweiten	113	17.48	der bilateralen Verhand-
					lungen
50	31.21	der im Weltkrieg	114	17.48	Millionen Franken im
51	26.22	Schweiz im Welt-	115	20.25	mit Millionen Fran-
		krieg			ken
52	38.65	der der Bundesrat	116	17.48	der Laufenden Rechnung
53	23.72	der der Bundesver-	117	17.48	überzeugt dass die
		fassung			
54	22.47	der der Materie	118	17.48	ist vor allem
55	22.47	der der Telecom	119	17.48	Bundesrat Jean-Pascal
					Delamuraz
56	15.24	der der Jahre	120	17.48	Bundesrat Flavio Cotti
57	23.72	der Bundesrat mit	121	16.23	zu Stimmen gutgeheissen
58	20.91	der Bundesrat der	122	24.97	wird zu Stimmen
59	18.73	der Bundesrat nicht	123	16.23	in Zusammenarbeit mit
60	26.22	der wegen der	124	16.23	Rolle der im
61	23.72	der Einheit der	125	47.44	die Rolle der
62	17.48	der Grund der	126	16.23	das Ergebnis der
63	16.23	der Mehrheit der	127	16.23	im Kampf gegen
64	29.96	Grund der der			

**Tabelle A.3:** Typische Mehrworteinheiten (Typikprofil) im Ressort ‚Inland‘ des NZZ-Korpus in den Jahren 1995–1997 im Vergleich zu den Jahren 2003–2005 (manuell gefiltert, Cluster,  $G^2 \geq 15$ ).

## A.2.2 Periode 2003–2005

Zeile	G <sup>2</sup>	MWE	Zeile	G <sup>2</sup>	MWE
1	−56.77	vor den Medien	82	−18.41	die Ausdehnung
2	−41.43	der der SVP	83	−19.95	neuen
3	−15.37	der der sei	84	−16.88	und die haben
4	−15.34	der der Kanton	85	−15.34	und die SP
5	−15.34	der der Bristen-	86	−19.95	FDP und die
		strasse			SVP und SP
6	−21.48	der SVP in	87	−19.95	Strom ohne Atom
7	−24.55	der gegenueber der	88	−19.95	sich in mit
8	−18.41	der Zahl der	89	−19.95	in von Millionen
9	−15.34	der Personenfrei-	90	−19.95	im Zuge der
		zuegigkeit der			
10	−15.34	der Bedeutung der	91	−19.95	folgt der Mehrheit
11	−15.34	der Armee der	92	−19.95	eine Reihe von
12	−16.88	der und SVP	93	−19.95	der Universitaet Basel
13	−19.95	Kritik der der	94	−30.69	fuer der Universi-
					taet
14	−15.34	Stimmen der der	95	−19.95	der St Gallen
15	−15.34	Jahren der der	96	−19.95	der Bund den
16	−24.35	die der SVP	97	−16.88	der des Bundesge-
					richts
17	−38.36	in den im	98	−15.34	der zwischen Bund
18	−35.29	in Bezug auf	99	−19.95	Bundesrat Hans-Rudolf
					Merz
19	−15.34	in Bezug die	100	−19.95	Bundespraesident
20	−33.76	Bundesrat Joseph Deiss	101	−18.41	Joseph Deiss
21	−31.99	der Schweiz an	102	−16.88	zu der Mehrheit
22	−16.88	der Schweiz haben	103	−18.41	zu der Rat
23	−15.34	einem der Schweiz	104	−15.34	wie der am
24	−30.69	die CVP die	105	−18.41	ebenso wie der
25	−21.48	die Ausdehnung	106	−18.41	von und Kantonen
		die			Stimmen der Mehrheit
26	−20.44	die FDP die	107	−15.34	Stimmen der Rat
27	−15.34	fuer die CVP	108	−18.41	sich fuer aus
28	−30.69	des Bundesamts fuer	109	−15.34	sich gegen aus
29	−29.15	von auf Prozent	110	−24.55	spricht sich aus
30	−29.15	Bund und Kantone	111	−18.41	Referendum gegen die
31	−24.75	Bund und Kanto-	112	−18.41	Position der NZZ
		nen			
32	−23.04	die Bund und	113	−18.41	Die Position NZZ
33	−21.73	von Bund und	114	−18.41	mit zu rechnen
34	−15.34	und Bund und	115	−18.41	Kanton St Gallen
35	−23.02	von Bund Kanto-	116	−18.41	Hoehe von Franken
		nen			
36	−16.88	der Bund Kanto-	117	−16.88	Hoehe von Millio-
		nen			nen
37	−29.15	AHV und IV	118	−18.41	Ausdehnung auf die
38	−15.34	der AHV IV	119	−21.48	die Ausdehnung
					auf

Tabelle A.4: (Fortsetzung folgende Seite)

**Tabelle A.4:** (Fortsetzung von vorheriger Seite)

Zeile	G <sup>2</sup>	MWE	Zeile	G <sup>2</sup>	MWE
39	-27.62	die SVP in	120	-18.41	Aus dem Versicherungsgericht
40	-18.41	die SVP den	121	-16.88	zur Bekaempfung der
41	-15.37	die SVP und	122	-16.88	wie sich die
42	-27	bei der nicht	123	-15.34	so wie die
43	-15.34	bei der Bristen- strasse	124	-15.34	ebenso wie die
44	-21.48	bei den nicht	125	-16.88	Pro Helvetia die
45	-26.09	in der Kantone	126	-16.88	in Zuerich und
46	-19.16	in der bis	127	-16.88	fuer Verteidigung und
47	-16.88	in der Verfassung	128	-20.44	Bundesamt fuer und
48	-15.34	in der seit	129	-16.88	Direktion fuer und
49	-15.34	Personen in der	130	-16.88	FDP und CVP
50	-26.09	der Vereinten Nationen	131	-16.88	die innere Sicherheit
51	-26.09	der auf neuen	132	-16.88	die embryonalen Stammzellen
52	-16.88	der neuen EU- Staaten	133	-16.88	die des Bundesgerichts
53	-15.34	Ausdehnung der auf	134	-16.88	der zweiten Saecule
54	-25.67	von bis zu	135	-16.88	der Steiner Chilbi
55	-25.67	mit den den	136	-16.88	der Bilateralen II
56	-24.55	und der EU	137	-16.88	der Armee XXI
57	-15.34	und der Armee	138	-16.88	der Armee Divisio- naer
58	-15.34	FDP und der	139	-23.02	Chef der Armee
59	-24.55	Die der EU	140	-16.88	Bundesraetin Calmy- Rey in
60	-18.41	Die der NZZ	141	-16.62	dass die Kantone
61	-24.55	Der der in	142	-16.62	10 vor 10
62	-24.55	Bundespraesident Pascal Couchepin	143	-15.34	zu Stimmen Rat
63	-24.55	an die werden	144	-15.34	Mit zu Rat
64	-16.62	an die Kantone	145	-15.34	von der FDP
65	-24.35	fuer die Jahre	146	-15.34	und in Bern
66	-16.88	Schweiz fuer die	147	-15.34	sich nicht nur
67	-23.44	der ueber das	148	-15.34	Schengen Dublin der
68	-15.34	der Bundesrat das	149	-15.34	Mit Stimmen Rat
69	-23.02	die von sind	150	-15.34	Mit der Rat
70	-19.95	die von Bund	151	-15.34	hat mit Stimmen
71	-21.48	die nicht sind	152	-15.34	gegen das der
72	-20.44	gegen die von	153	-15.34	die Wirtschaft und
73	-23.02	dass der sei	154	-15.34	die nach einem
74	-23.02	CVP und FDP	155	-15.34	der von Bern
75	-21.48	Personenfreizuegigkeit auf die	156	-15.34	der flankierenden Mass- nahmen
76	-15.34	Ausdehnung Personenfreizuegigkeit auf	157	-15.34	der Eroeffnung des
77	-32.22	der Personenfrei- zuegigkeit die	158	-15.34	Bundesraetin Ruth Metzler

**Tabelle A.4:** (Fortsetzung folgende Seite)

**Tabelle A.4:** (Fortsetzung von vorheriger Seite)

Zeile	G <sup>2</sup>	MWE	Zeile	G <sup>2</sup>	MWE
78	−16.88	Ausdehnung Personenfreizuegigkeit die	159	−15.34	7 Millionen Franken
79	−21.48	in Schweiz zu	160	−15.21	im vergangenen Jahr
80	−21.48	die neuen EU-Staaten	161	−31.89	der im Jahr
81	−21.48	die auf neuen			

**Tabelle A.4:** Typische Mehrworteinheiten (Typikprofil) im Ressort ‚Inland‘ des NZZ-Korpus in den Jahren 2003–2005 im Vergleich zu den Jahren 1995–1997 (manuell gefiltert, Cluster,  $G^2 \geq 15$ ).

### A.3 Ressort ‚Feuilleton‘

#### A.3.1 Periode 1995–1997

Zeile	G <sup>2</sup>	MWE	Zeile	G <sup>2</sup>	MWE
1	303.21	schweigen schweigen schweigen	42	21.82	eine aus dem
2	17.23	schweigen schweigen nur	43	21.82	in der deutschen
3	47.09	Hinweise auf Buecher	44	18.36	in der werden
4	35.60	der Kunst der	45	20.67	wie auch die
5	22.00	der Welt der	46	20.67	zur und der
6	20.67	der wird der	47	20.67	Das ist und
7	18.38	von der Kunst	48	18.38	Das ist nicht
8	20.48	zwischen der der	49	19.52	aus der Sammlung
9	16.83	hat der der	50	19.52	und zu sein
10	15.03	nach der der	51	19.52	und den nicht
11	27.56	der eine von	52	17.23	und noch nicht
12	20.67	der sich eine	53	19.32	der New York
13	17.23	der als einer	54	18.53	des in und
14	17.15	Die der von	55	18.38	der des ein
15	26.42	die noch die	56	16.47	der des den
16	26.42	aus den und	57	18.38	wird der des
17	26.42	das sich und	58	16.08	In der ein
18	26.42	wie das der	59	18.38	die sich an
19	25.27	die der werden	60	15.35	die sich als
20	16.08	die der Kuenstler	61	18.38	laesst sich in
21	16.82	die als der	62	18.26	unter Leitung von
22	25.27	dass er die	63	17.23	die Welt in
23	16.08	dass sie die	64	16.08	hat die in
24	16.00	dass der die	65	17.23	haben die und
25	16.08	der dass die	66	17.23	und und wieder
26	25.27	der Jahre die	67	15.85	und und mit
27	17.41	einer der die	68	16.28	und zwischen und
28	25.16	er sich mit	69	26.42	zu und und
29	24.12	der beiden und	70	17.23	zwischen und zwischen
30	18.38	Geschichte der und	71	17.23	die von bis
31	22.97	wird in die	72	16.47	auf den die
32	16.08	wird der in	73	16.45	als eine der
33	22.97	dass es nicht	74	16.08	nicht mehr und
34	22.97	von ueber und	75	16.08	in den ist
35	19.39	des von und	76	16.08	zu Beginn der
36	22.97	auf Grund der	77	16.08	der fruehen Neuzeit
37	22.22	und die eine	78	16.08	wird von den
38	21.82	und auch auf	79	16.08	mit einem zu
39	16.08	und auf von	80	24.34	die mit einem
40	17.23	und damit auch			

**Tabelle A.5:** Typische Mehrworteinheiten (Typikprofil) im Ressort ‚Feuilleton‘ des NZZ-Korpus in den Jahren 1995–1997 im Vergleich zu den Jahren 2003–2005 (manuell gefiltert, Cluster,  $G^2 \geq 15$ ).

## A.3.2 Periode 2003–2005

Zeile	G <sup>2</sup>	MWE	Zeile	G <sup>2</sup>	MWE
1	−48.03	die das des	48	−19.87	Duke of Portland
2	−36.44	zum ersten Mal	49	−19.87	als und des
3	−31.47	mit der er	50	−19.87	aber auch nicht
4	−29.81	es nicht mehr	51	−19.03	In den Jahren
5	−21.53	und es nicht	52	−18.22	mit einer zu
6	−28.16	der mit seinen	53	−18.22	Frankfurt Main 2003
7	−26.5	und ihre in	54	−18.22	Frankfurt am 2003
8	−18.71	und in Zuerich	55	−18.22	die von aus
9	−28.16	ein und in	56	−29.81	als die von
10	−26.5	Johannes Paul II	57	−18.22	der Paris Review
11	−26.5	der ist es	58	−18.22	den USA und
12	−18.76	der nicht ist	59	−27.3	in den USA
13	−24.84	The of the	60	−18.22	den den von
14	−24.84	dem in und	61	−18.22	sich den von
15	−23.19	ueber die bis	62	−18.22	aus und den
16	−23.19	der Gesellschaft und	63	−18.22	am Main 2003
17	−16.56	der Frau und	64	−17.39	sich und als
18	−23.19	dass er in	65	−16.56	Weite See Die
19	−22.85	die fuer den	66	−23.19	Die Weite See
20	−21.53	von das der	67	−16.56	war und der
21	−16.56	von der Musik	68	−16.56	Verlag Nagel Kimche
22	−16.56	von der als	69	−16.56	sich wie ein
23	−21.53	ist im der	70	−16.56	nicht zur Existenz
24	−21.53	ist das in	71	−16.56	nicht fuer die
25	−16.56	ist in eine	72	−16.56	in einer Ausstellung
26	−21.53	ein mit einem	73	−16.56	in der Erzählung
27	−21.53	Die Binnensee Die	74	−39.75	ueber in der
28	−16.56	Die Die Binnensee	75	−21.53	Zuercher in der
29	−21.53	der Universitaet Zuerich	76	−21.53	was in der
30	−21.53	dass der ein	77	−18.22	Zuerich in der
31	−21.53	das Lucerne Festival	78	−16.56	fuer und des
32	−21.53	Blick auf und	79	−21.53	auch fuer und
33	−21.53	an die an	80	−16.56	die zu bringen
34	−16.56	man sich an	81	−21.53	und die zur
35	−23.19	das an die	82	−16.56	die in Europa
36	−20.6	die oder die	83	−18.76	den die in
37	−20.14	der von von	84	−16.56	der Stadt zu
38	−19.87	vom und von	85	−19.87	der der zum
39	−19.87	ueber das von	86	−16.56	der noch in
40	−19.87	The Free Design	87	−16.56	der fuer eine
41	−19.87	sie auch die	88	−16.56	aus der aus
42	−19.87	sich der aus	89	−23.19	auch aus der
43	−19.87	New Europe College	90	−16.56	auf dem von
44	−19.87	ist ein fuer	91	−15.55	die der im
45	−16.56	Die ist ein	92	−15.41	Die die der
46	−19.87	ist die Geschichte			

**Tabelle A.6:** Typische Mehrworteinheiten (Typikprofil) im Ressort ‚Feuilleton‘ des NZZ-Korpus in den Jahren 2003–2005 im Vergleich zu den Jahren 1995–1997 (manuell gefiltert, Cluster,  $G^2 \geq 15$ ).

## A.4 Ressort ‚Wirtschaft‘

### A.4.1 Periode 1995–1997

Zeile	G <sup>2</sup>	MWE	Zeile	G <sup>2</sup>	MWE
1	181.88	16 Uhr mit	69	60.22	der achtziger Jahre
2	86.03	16 Uhr bei	70	54.07	Primaermarktkurse
3	459.63	um 16 Uhr	71	54.07	Emissions- Rendite
4	176.97	Dollar 16 Uhr	72	54.07	Primaermarktkurse
5	111.83	Zuerich 16 Uhr	73	54.07	fuer Emissions-
6	99.54	lag 16 Uhr	74	54.07	Primaermarktkurse
7	88.48	sich 16 Uhr	75	54.07	Franken-Neuemissionen
8	84.80	dreimonatige 16 Uhr	76	54.07	Emissions-
9	67.59	tiefer 16 Uhr	77	34.41	Primaermarktkurse
10	31.95	staerker 16 Uhr	78	54.07	Betrag Emissions-
11	168.37	lag um bei	79	54.07	Primaermarktkurse
12	36.87	lag der um	80	54.07	fuer Emissions-
13	164.68	lag Uhr bei	81	54.07	Franken-Neuemissionen
14	78.65	lag 16 bei	82	54.07	Emissions-
15	44.24	lag der bei	83	52.84	Primaermarktkurse
16	30.72	und lag bei	84	52.09	fuer Rendite
17	140.10	als am Vortag	85	51.62	Primaermarktkurse
18	36.87	als am Freitag	86	46.70	fuer Betrag
19	121.67	Basispunkte als am	87	35.64	Primaermarktkurse
20	62.68	hoeher als am	88	111.83	Franken-Neuemissionen
21	58.99	tiefer als am	89	33.18	Betrag
22	100.77	um als Vortag	90	46.70	Schweiz Primaer-
23	79.88	Basispunkte als	91	35.34	marktkurse fuer
24	67.59	Vortag	92	46.41	Bank von England
25	44.24	damit als Vortag	93	45.88	in einer der
26	36.87	tiefer als Vortag	94	41.78	France OAT Dom
27	120.44	hoeher als Vortag	95	45.47	Basispunkte tiefer als
28	103.23	in in Lokalwaehrung	96	174.05	Basispunkte hoeher
		in in Osten			als
					um Basispunkte als
					und Basispunkte als
					teilte weiter mit
					teilte das mit
					die des Landes
					sich der fuer
					stellte sich der
					Grund der der
					auf Grund der

**Tabelle A.7:** (Fortsetzung folgende Seite)

Tabelle A.7: (Fortsetzung von vorheriger Seite)

Zeile	G <sup>2</sup>	MWE	Zeile	G <sup>2</sup>	MWE
29	103.23	in in Mittlerer	97	45.47	tiefer am Vortag
30	89.93	in in Europa	98	45.47	die Fusion der
31	103.23	in Lokalwaehrung	99	38.10	die zwischen der
32	103.23	Europa in Lokalwaehrung	100	34.31	die Regierung der
33	103.23	Osten in Lokalwaehrung	101	39.33	etwa die der
34	68.82	Mittlerer in Lokalwaehrung in	102	44.24	Boerse hat geschlossen
35	39.33	in Schweiz in	103	65.22	Die Boerse hat
36	100.77	um am Vortag	104	41.78	ergab sich um
37	30.72	um am Freitag	105	33.18	ergab sich Uhr
38	55.30	um tiefer am	106	41.78	Durchschnittsrendite der Bundesobligationen
39	50.39	um hoeher am	107	52.84	Die Durchschnitts- rendite der
40	41.78	um tiefer Vortag	108	41.78	der von England
41	97.09	dem Vortagesschluss von	109	39.33	in Zuerich um
42	33.18	dem Wert von	110	36.87	Geld- und Kapitalmarkt
43	345.16	gegenueber dem von	111	36.87	hoeher am Vortag
44	79.88	Basispunkte am Vortag	112	35.64	einem an die
45	46.70	Basispunkte tiefer	113	35.64	Dollar Zuerich um
46	35.64	am Basispunkte hoeher	114	34.41	wurde die mit
47	35.64	am Basispunkte tiefer	115	34.41	den 9 Monaten
48	79.88	Vortag damit Basispunkte	116	34.41	Auslaendische Unterneh- men der
49	33.18	am und Basispunkte am	117	33.79	der Europaeischen Union
50	76.19	fuer dreimonatige Uhr	118	33.18	in Conf-Futures Kontrak- te
51	72.51	fuer dreimonatige	119	57.76	jener in Conf- Futures
52	57.76	um Satz fuer dreimonati- ge	120	33.18	Electric Cap Corp
53	73.74	des US-Schatzamtes dem	121	38.10	General Electric Corp
54	72.51	des US-Schatzamtes gegenueber	122	33.18	in bezug auf
55	82.34	Staaten des US- Schatzamtes	123	33.18	dass die Regierung
56	72.51	Rendite des US- Schatzamtes	124	33.18	ersten 9 Monaten
57	68.82	Die des US- Schatzamtes	125	31.95	Long Bond US- Schatzamtes
58	31.95	Bond des US- Schatzamtes	126	31.95	Long Bond des
59	31.95	Long des US- Schatzamtes	127	31.95	der Bank Baer
60	63.90	Rendite des dem	128	71.28	der Schweizerischen Bankgesellschaft
61	70.05	in Europa Osten	129	31.95	staerker Zuerich Uhr

Tabelle A.7: (Fortsetzung folgende Seite)



**Tabelle A.7:** (Fortsetzung von vorheriger Seite)

Zeile	G <sup>2</sup>	MWE	Zeile	G <sup>2</sup>	MWE
62	68.82	in Europa Mittlerer	130	31.95	staerker Zuerich um
63	68.82	in Mittlerer Osten	131	31.95	Betrag in Mio
64	66.36	der CS Holding	132	31.23	von der Regierung
65	60.22	der Holding AG	133	30.72	Die deutschen im
66	63.90	Rendite des gegenueber	134	30.72	der Schweizer Rueck
67	68.82	Vereinigte Rendite des			

**Tabelle A.7:** Typische Mehrworteinheiten (Typikprofil) im Ressort ‚Wirtschaft‘ des NZZ-Korpus in den Jahren 1995–1997 im Vergleich zu den Jahren 2003–2005 (manuell gefiltert, Cluster,  $G^2 \geq 15$ ).

## A.4.2 Periode 2003–2005

Zeile	G <sup>2</sup>	MWE	Zeile	G <sup>2</sup>	MWE
1	−87.2	Amortisation Amortisation	61	−31.14	In stieg um
2	−71.63	Amortisation Amortisation Investition	62	−46.71	ersten neun Monaten
3	−60.73	Amortisation Amortisation und	63	−31.66	den ersten neun
4	−42.04	Amortisation Amortisation Betrieb	64	−46.71	Die am Montag
5	−43.6	Heizöl Amortisation Amortisation	65	−109.76	Die haben am
6	−82.78	Die Aktien von	66	−46.71	Der Preis von
7	−73.18	Die Aktien des	67	−45.16	das Unternehmen der
8	−51.38	Die Aktien um	68	−43.6	Vortrag Veraenderung in
9	−31.28	Die Aktien der	69	−43.6	Vortrag Veraenderung Franken
10	−54.5	Die Papiere von	70	−43.6	Vortrag Aktuell in
11	−49.83	Die Titel von	71	−43.6	Uhr Vortrag Veraenderung
12	−80.97	mit Plus von	72	−43.6	Uhr Vortrag Franken
13	−79.41	mit einem Plus	73	−43.6	Uhr Vortrag Aktuell
14	−76.3	die Aktien des	74	−43.6	Uhr Vortrag 1
15	−66.6	die Aktien von	75	−43.6	Uhr Veraenderung Franken
16	−52.94	die Titel des	76	−43.6	Uhr Aktuell Franken
17	−38.93	die Papiere des	77	−43.6	Uhr Veraenderung in
18	−73.18	schloss der um	78	−43.6	Marktuebersicht Devisen Asien
19	−105.88	In schloss der	79	−35.81	Marktuebersicht Devisen Dollar
20	−56.06	In schloss um	80	−32.7	Marktuebersicht Devisen Europa
21	−68.84	an die Boerse	81	−43.6	Aktuell Veraenderung in
22	−68.51	der Deutschen Boerse	82	−43.6	Aktuell Veraenderung Franken
23	−38.93	der Boerse SWX	83	−42.22	hat im ersten
24	−66.96	In der von	84	−42.04	an den Aktienmaerkten
25	−42.04	In der Hang-Seng-Index	85	−38.93	Der Dow Jones
26	−46.71	In Hongkong der	86	−38.93	Cent in Yen
27	−40.48	In stieg der	87	−38.93	Cent in Dollar
28	−34.26	In London der	88	−38.69	Geld und Anlage
29	−32.7	In Frankfurt der	89	−37.37	Zu den zaehlten

Tabelle A.8: (Fortsetzung folgende Seite)

**Tabelle A.8:** (Fortsetzung von vorheriger Seite)

Zeile	G <sup>2</sup>	MWE	Zeile	G <sup>2</sup>	MWE
30	-63.74	im dritten Quartal	90	-37.37	haben am Montag
31	-60.73	im Quartal von	91	-37.37	der standen die
32	-38.93	im Quartal einen	92	-32.7	der Anleger die
33	-37.19	im Quartal um	93	-37.37	Aktie verglichen mit
34	-46.57	im ersten Quar-	94	-40.48	je Aktie mit
		tal			
35	-30.55	im vierten Quar-	95	-35.81	Jahr einen von
		tal			
36	-31.14	hat im dritten	96	-35.81	der der Euro-Zone
37	-105.41	hat im Quartal	97	-34.82	den vergangenen
					Jahren
38	-60.59	fuer das Quartal	98	-42.01	in den vergange-
					nen
39	-56.06	im Jahr 2004	99	-34.42	Quartal einen von
40	-49.83	im Jahr 2003	100	-34.42	Kennzahlen in Mio
41	-52.94	und Betrieb und	101	-34.26	Leistungs- Struktur-
					daten Mio
42	-52.94	und Betrieb	102	-34.26	In den USA
		Investition			
43	-62.28	Investition und	103	-34.26	im gleichen Vorjahres-
		Betrieb			abschnitt
44	-52.94	Betrieb und	104	-35.81	oder im gleichen
		Betrieb			
45	-52.94	In schloss o	105	-35.81	Cent im gleichen
46	-52.94	in Franken 2	106	-34.26	der Hang-Seng-Index
					o
47	-43.6	Vortrag in Fran-	107	-42.04	Hongkong der
		ken			Hang-Seng-Index
48	-43.6	Veraenderung in	108	-33.04	Boersen und der
		Franken			
49	-43.6	Uhr in Franken	109	-32.7	Der notierte bei
50	-43.6	Aktuell in Fran-	110	-31.14	Titel von um
		ken			
51	-118.34	Yen in Franken	111	-49.83	Die Titel um
52	-51.38	und Maerkte Euro	112	-31.14	Rendite Risiko Ratio
53	-34.42	und Maerkte der	113	-31.14	Information
					Rendite Risiko
54	-43.6	Konjunktur und	114	-31.14	Information
		Maerkte			Rendite Ratio
55	-49.83	Boersen und	115	-31.14	Erwartungen der
		Euro			Analytiker
56	-49.83	Schweizer Boerse	116	-31.14	Cent je verglichen
		SWX			
57	-32.7	der Schweizer	117	-73.18	oder Cent je
		SWX			
58	-49.83	Boersen Maerkte	118	-31.14	Bank of America
		Euro			
59	-46.95	stieg der um	119	-109	einem Plus von
60	-42.04	stieg um 1			

**Tabelle A.8:** Typische Mehrworteinheiten (Typikprofil) im Ressort 'Wirtschaft' des NZZ-Korpus in den Jahren 2003-2005 im Vergleich zu den Jahren 1995-1997 (manuell gefiltert, Cluster,  $G^2 \geq 15$ ).

## A.5 Ressort ‚Sport‘

## A.5.1 Periode 1995–1997

Zeile	G <sup>2</sup>	MWE	Zeile	G <sup>2</sup>	MWE
1	79.40	Die Spiele der	35	19.85	sind in der
2	27.29	Die Spiele vom	36	18.55	gegen in der
3	24.81	Die Spiele naechsten	37	23.57	und sich mit
4	24.81	Die Spiele Runde	38	22.33	Die naechsten Runde
5	41.53	Die der der	39	22.33	in auch in
6	35.98	Die der Runde	40	19.85	in Sierra Nevada
7	35.98	Die der naechsten	41	19.85	durch die und
8	18.61	Die Runde der	42	19.85	GP von Belgien
9	16.13	Die Schweizer der	43	19.85	der gegen in
10	34.74	die des die	44	16.13	der in Atlanta
11	34.74	der Sierra Nevada	45	18.61	der Damen in
12	17.37	in der Sierra	46	20.70	als der in
13	17.37	in der Nevada	47	18.61	der Hockey League
14	34.74	ueber 5000 m	48	18.61	km vor dem
15	19.85	ueber 3000 m	49	17.37	km vor Ziel
16	33.33	nach wie vor	50	17.37	km dem Ziel
17	27.29	um 20 Uhr	51	18.61	in vergangenen Jahren
18	26.05	der Schweizer im	52	17.37	Los Angeles Kings
19	26.05	Die im Telegramm	53	17.37	New York Rangers
20	26.05	gegen den SC	54	17.37	Matt in Zuegen
21	26.05	der Runde der	55	17.37	und U1 Gruppe
22	20.70	der naechsten Runde	56	16.13	der den Jahren
23	20.96	in der Runde	57	17.37	ist der den
24	17.37	Spiele der Runde	58	16.13	in Val-d Isère
25	26.05	sich der nicht	59	16.13	mal 2 Minuten
26	23.57	die sich der	60	43.42	5mal 2 Minuten
27	24.81	in mehr als	61	24.81	6mal 2 Minuten
28	24.81	der Copa America	62	21.09	4mal 2 Minuten
29	24.81	und aus der	63	16.13	In der das
30	26.05	an und der	64	16.13	um Platz 3
31	24.81	Fakten und Facetten	65	18.62	Spiel um Platz
32	23.57	in der sein	66	16.13	Herren um Platz
33	23.57	in der NHL			

**Tabelle A.9:** Typische Mehrworteinheiten (Typikprofil) im Ressort ‚Sport‘ des NZZ-Korpus in den Jahren 1995–1997 im Vergleich zu den Jahren 2003–2005 (manuell gefiltert, Cluster,  $G^2 \geq 15$ ).

## A.5.2 Periode 2003–2005

Zeile	G <sup>2</sup>	MWE	Zeile	G <sup>2</sup>	MWE
1	−64.07	ooo Zuschauer Tore	47	−18.52	2 Minuten Lugano
2	−15.52	ooo Zuschauer Tor	48	−16.98	2 Minuten 5-mal
3	−46.3	von Swiss Olympic	49	−74.08	5-mal 2 Minuten
4	−26.24	von Swiss Ski	50	−66.36	6-mal 2 Minuten
5	−21.61	der von Swiss	51	−58.65	4-mal 2 Minuten
6	−40.91	Rangliste je Spiele	52	−44.76	7-mal 2 Minuten
7	−37.04	Rote Karte gegen	53	−29.32	8-mal 2 Minuten
8	−33.95	zum Mal in	54	−27.95	gegen 2 Minuten
9	−23.15	zum ersten Mal	55	−21.61	9-mal 2 Minuten
10	−21.61	zum zweiten Mal	56	−16.98	3-mal 2 Minuten
11	−21.61	zum dritten Mal	57	−18.05	Los Angeles Lakers
12	−27.78	Runde mit Schweiz	58	−16.98	Stand nach dem
13	−27.78	in gegen den	59	−16.98	Stade de Genève
14	−26.24	m Crawl m	60	−16.98	Sampdoria Genua 1
15	−18.52	m Ruecken m	61	−16.98	Jan von Arx
16	−18.52	m 200 m	62	−16.98	Golden State Warriors
17	−16.98	ueber m Crawl	63	−16.98	die AS Roma
18	−18.52	200 m m	64	−16.58	ins leere Tor
19	−23.15	WTA-Turnier ooo Dollar	65	−15.43	ZSC Lions 3
20	−23.15	alle verletzt und	66	−15.43	Wort zum Sport
21	−28.62	und alle verletzt	67	−15.43	und im Super-G
22	−22.79	vor zwei Jahren	68	−15.43	Swiss Football League
23	−21.93	den Spielen in	69	−30.87	der Swiss League
24	−21.61	Minuten gegen 6-mal	70	−15.43	Roter Stern Belgrad
25	−18.52	Minuten gegen Lugano	71	−15.43	rote Karte gegen
26	−16.98	Minuten gegen 5-mal	72	−30.87	Gelb-rote Karte gegen
27	−21.61	in die Serie	73	−15.43	Right to Play
28	−16.98	in Serie B	74	−15.43	New York Major
29	−21.61	ersten in der	75	−15.43	Moskau 1 1
30	−21.61	der Serie B	76	−15.43	in diesem Winter
31	−17.97	der Serie A	77	−15.43	hatte er in
32	−26.24	der FC Basel	78	−15.43	GCK Lions 2
33	−20.14	Reto von Arx	79	−15.43	er mit den
34	−20.06	West Bromwich Albion	80	−20.06	der Saison mit
35	−20.06	Marazzi De Maria	81	−15.43	er an der
36	−20.06	Kenteris und Thanou	82	−16.98	der der Tour
37	−20.06	der Regular Season	83	−15.43	die der WM
38	−20.06	Der FC Thun	84	−15.43	der Partie gegen
39	−16.98	Der FC gegen	85	−15.43	der im nicht
40	−18.97	noch in der	86	−15.43	der im Jahr
41	−18.52	in La Chaux-de-Fonds	87	−15.43	an Olympischen in
42	−18.52	er sich die	88	−15.43	An den die
43	−18.52	die in Schweiz	89	−15.43	60 m Huerden
44	−24.69	mit die in	90	−15.43	200 m Lagen
45	−18.52	der Frauen in	91	−15.43	100 m Lagen
46	−18.52	50 m Brust			

**Tabelle A.10:** Typische Mehrworteinheiten (Typikprofil) im Ressort ‚Sport‘ des NZZ-Korpus in den Jahren 2003–2005 im Vergleich zu den Jahren 1995–1997 (manuell gefiltert, Cluster,  $G^2 \geq 15$ ).

## A.6 Ressort ‚Leserbriefe‘

In der folgenden Tabelle sind die typischsten Mehrworteinheiten im Ressort ‚Leserbriefe‘ im Vergleich zu allen anderen Ressorts über die ganze Zeitperiode 1995–2005 hinweg dargestellt.

Zeile	G <sup>2</sup>	MWE	Zeile	G <sup>2</sup>	MWE
1	261.30	In der vom	53	50.95	Zum von Stuerm
2	217.18	Es ist dass	54	59.44	Galadiner JWC in
3	109.07	Es ist zu	55	59.44	Galadiner des in
4	59.44	Es ist fuer	56	59.44	die noch viel
5	54.56	Es ist nicht	57	59.44	in seinem Artikel
6	101.90	Es zu dass	58	59.44	Ich bin dass
7	107.81	an die nicht	59	59.44	des JWC in
8	93.41	an die ich	60	59.44	die Eigenschaften Teile
9	67.93	an die Beduerfnisse	61	59.44	Reisesteuer zur Neat-Finanzierung
10	67.93	an die vom	62	59.44	Tod Walter Stuerm
11	63.43	an die im	63	50.95	Tod von Stuerm
12	60.88	an die von	64	59.44	ist wieder Regensburg
13	59.08	an die des	65	59.44	die Schuld der
14	104.27	die Schweiz und	66	59.44	die Gesundheit der
15	59.44	Europaeische die	67	59.44	Betteln in Zuerich
16	59.44	Schweiz	68	59.44	Taeglich 15 Millionen
17	76.42	Massarbeit die	69	59.44	Wirtschaftliche Leichenschaendung der
18	50.95	Schweiz	70	59.44	von Walter Stuerm
19	76.42	Goeteborg oder der	71	58.41	ist der Fall
20	67.93	Goeteborg der Naivitaet	72	52.44	ist nicht der
21	69.90	Autos in Japan	73	56.42	dass auf den
22	67.93	Amerikanische	74	54.53	wenn sie nicht
23	50.95	Autos Japan	75	52.77	Streit zwischen und
24	67.93	nichts zu tun	76	51.96	und nicht mit
25	67.93	Regensburg wieder Regensburg	77	50.95	und ihre Geschichte
26	59.44	Regensburg ist wieder	78	50.95	Achzen im Bau
27	67.93	Amerikanische in Japan	79	50.95	kann man die
28	67.93	der der Naivitaet	80	50.95	Toetung von Kueken
29	67.93	oder Preis der	81	50.95	Arzneiversand der Krankenkassen
30	67.93	aus dem Bundesamt	82	50.95	stirbt die Stadt
31	67.93	es sei denn	83	59.44	Bauern stirbt die
32	67.93	hat das Recht	84	50.95	Ohne stirbt Stadt
33	67.93	Es hat Recht	85	50.95	Spuren des Kulturkampfes
		fuer den Geburtenrueckgang			
		fuer Kofi Annan			
		Friedensnobelpreis			
		fuer Kofi			

Tabelle A.11: (Fortsetzung nächste Seite)

**Tabelle A.11:** (Fortsetzung von vorheriger Seite)

Zeile	G <sup>2</sup>	MWE	Zeile	G <sup>2</sup>	MWE
34	67.93	und der Juden	86	50.95	der habe ich
35	59.44	und der Eizenstat- Bericht	87	50.95	Islam heiliger Krieg
36	53.18	und der Schweiz	88	50.95	Freilandhaltung statt Toetung
37	67.93	Wir und der	89	50.95	im Wuergegriff Israels
38	50.95	Jugoslawien und der	90	50.95	Fifa in Zuerich
39	67.93	Machtkampfen bedrohte Aphrodite	91	50.95	die Nichts gelernt
40	67.93	nicht immer recht	92	50.95	Schweiz in Krise
41	50.95	hat nicht immer	93	50.95	Volk hat immer
42	67.27	in der Schweiz	94	59.44	Das Volk hat
43	51.26	in der schon	95	50.95	Aus der Bankangestellten
44	50.95	in der Kuestenab- schnitte	96	50.95	Unsinn dem Bundesamt
45	118.88	Artikel in der	97	50.95	Zwei Gesichter des
46	63.27	wir in der	98	50.95	Eichel die EU
47	67.23	dass die ist	99	50.95	ich ueber die
48	58.10	dass es ist	100	50.95	die von Herrn
49	59.44	hoffen dass die	101	50.95	Tatbeweis statt Kriegs- handwerk
50	60.49	zu hoffen dass	102	50.95	Hotelbadezimmer allen Schikanen
51	59.44	Zum Walter Stuerm	103	50.95	Europaeische fuer die
52	50.95	Zum Tod Walter			

**Tabelle A.11:** Typische Mehrworteinheiten im Ressort ‚Leserbriefe‘ des NZZ-Korpus in den Jahren 1995–2005 im Vergleich zu den anderen Ressorts (manuell gefiltert, Top 103).

## Tabellenverzeichnis

4.1	Experiment ‚Textroutinen‘ von Feilke. . . . .	64
6.1	Grundsätzliche Möglichkeiten der Berechnung von Kollokationen. . . . .	117
6.2	Ausschnitt aus der Kookkurrenzdatenbank CCDB (Belica 2001–2006) zum Analysewort <i>Zahl</i> . . . . .	120
6.3	Kollokationen von <i>facing</i> und <i>faced</i> im ‚Birmingham Corpus‘ und im ‚Economist and Wall Street Journal Corpus‘. . . . .	125
7.1	Kontingenztafel für zwei Variablen S und P. . . . .	136
7.2	Der Mann-Whitney-Rank-Test für zwei Korpora X und Y. . . . .	140
7.3	Berechnung des Dispersionsmaßes Juillands D für einige Einzelwörter und Kollokationen im NZZ-Korpus. . . . .	143
7.4	Statistische Maße und ihre Stärken und Schwächen hinsichtlich korpuslinguistischer Funktionen. . . . .	145
8.1	Methoden der Interpretation . . . . .	155
8.2	Kontingenztafeln für <i>Kampf gegen</i> im NZZ-Korpus (Basis: Artikel). . . . .	157
8.3	Statistik von <i>Kampf gegen (den) Terror</i> im NZZ-Korpus. . . . .	158
8.4	$\chi^2$ -Statistik (Einheit: Artikel) für <i>in den Tod</i> in Artikeln des Ressorts ‚Ausland‘ im NZZ-Korpus. . . . .	164
8.5	$\chi^2$ -Statistik zur Verwendung von Ausdrücken des Paradigmas STERBEN in der Auslandsberichterstattung des NZZ-Korpus. . . . .	167
9.1	Die wichtigsten öffentlich verfügbaren deutschsprachigen Korpora im Vergleich. . . . .	179
9.2	Eine Auswahl an Software zur corpus-driven Berechnung von Kollokationen. . . . .	183
10.1	Beglaubigte Auflagen der ‚Neuen Zürcher Zeitung‘ . . . . .	190



10.2	Datengrundlage der Untersuchung. Artikel der Neuen Zürcher Zeitung zwischen 1995 und 2005. . . . .	191
11.1	Ressorts im NZZ-Korpus. . . . .	195
12.1	Typische Mehrworteinheiten im Ressort ‚Ausland‘ des NZZ-Korpus in den Jahren 1995–1997 im Vergleich zu den Jahren 2003–2005. . . . .	201
12.2	Typische Mehrworteinheiten im Ressort ‚Ausland‘ des NZZ-Korpus in den Jahren 2003–2005 im Vergleich zu den Jahren 1995–1997. . . . .	202
13.1	Typische Mehrworteinheiten im Ressort ‚Ausland‘ in den Zeitperioden 1995–1997 und 2003–2005. . . . .	210
13.2	Verteilung und $\chi^2$ -Statistik (Einheit: Artikel) von <i>bosnische(n) Serben</i> im NZZ-Korpus. . . . .	213
13.4	Die Vorkommen von <i>die/den Deutschen, deutsch</i> und <i>Deutschland</i> in den Ressorts des NZZ-Korpus. . . . .	218
13.5	Verteilungen und $\chi^2$ -Statistiken (Einheit: Artikel) für die Mehrworteinheiten <i>Kampf(s/es) gegen ... Terror*</i> bzw. <i>Kampf(s/es) gegen</i> im NZZ-Korpus. . . . .	221
13.6	Die Füllungen von X in den Mustern KAMPF MIT X und KAMPF DEM/DER [DATIVOBJEKT] im NZZ-Korpus. . . . .	224
13.7	Die Füllungen von X in den Mustern KAMPF GEGEN X und KAMPF MIT X im IDS-Korpus. . . . .	225
13.8	Die Füllungen von X im Muster KAMPF DEM X im IDS-Korpus. . . . .	226
13.9	Verteilungen und $\chi^2$ -Statistiken (Einheit: Artikel) von <i>die Zahl der</i> im ganzen Korpus und in Artikeln des Ressorts ‚Ausland‘. . . . .	233
13.10	Verteilung der Artikel mit mindestens einem Vorkommen von <i>die Zahl der</i> auf die Ressorts vor und nach dem Jahr 2000. . . . .	233
13.11	Kontexte von <i>die Zahl der X</i> vor und nach dem Jahr 2000 im Ressort ‚Ausland‘. . . . .	234
13.12	X der Wendung <i>die Zahl der X</i> vor und nach dem Jahr 2000 für die Kategorien KRIEG, KRIMINALITÄT und POLITIK im Ressort ‚Ausland‘. . . . .	235

13.13	Frequenzen für X in Wendungen <i>[Zahl] ... X</i> , wobei X ein Nomen und die Zahl weder Ordinal- noch Jahreszahl ist. . . . .	237
13.14	Verteilungen und $\chi^2$ -Statistiken (Einheit: Artikel) von <i>[Zahl] Mann</i> bzw. <i>[Zahl] Soldaten</i> in Artikeln des Ressorts ‚Ausland‘. . . . .	240
13.15	Typische Mehrworteinheiten im Ressort ‚Inland‘ in den Zeitperioden 1995–1997 und 2003–2005. . . . .	241
13.16	Verteilungen und $\chi^2$ -Statistiken (Einheit: Artikel) von <i>in den Medien</i> in Artikeln des Ressorts ‚Inland‘ und in allen Ressorts. . . . .	244
13.17	Verteilungen (Einheit: Artikel) von <i>in den Medien</i> im ‚St. Galler Tagblatt‘ und im ‚Zürcher Tages-Anzeiger‘ auf Basis des DeReKo IDS (o. J.)-Korpus. . . . .	245
13.18	Signifikante Mehrworteinheiten mit dem Bestandteil <i>Medien</i> im Ressort ‚Inland‘ in den Zeitperioden 1995–1997 und 2003–2005. . . . .	247
13.19	Die X in den Mustern KAMPF GEGEN ... X bzw. BEKÄMPFUNG ... X mit Mindestfrequenz 2 im Ressort ‚Inland‘. . . . .	249
13.20	Die X in den Mustern KAMPF GEGEN ... X bzw. BEKÄMPFUNG ... X mit Mindestfrequenz 3 im Amtlichen Bulletin des Schweizer Parlaments 1999–2003. . . . .	251
13.21	‚Probleme der Schweiz‘ gemäß Umfrage des ‚Sorgenbarometers‘ 1999–2005. . . . .	253
13.22	Typische Mehrworteinheiten aus dem Ressort ‚Inland‘ in den Zeitperioden 1995–1997 (links) und 2003–2005 (rechts) mit dem Ausdruck <i>nicht nur</i> als Bestandteil. . . . .	257
13.23	Verteilungen und $\chi^2$ -Statistiken (Einheit: Artikel) von <i>sich ... nicht nur</i> und <i>nicht nur ... sondern (... auch)</i> in Artikeln des Ressorts ‚Inland‘. . . . .	259
13.24	Verteilungen und $\chi^2$ -Statistiken (Einheit: Artikel) von <i>nicht nur ... sondern</i> über die Ressorts des NZZ-Korpus. . . . .	261
13.25	Verteilungen (Einheit: Artikel) von <i>nicht nur ... sondern</i> über die Originalressorts des NZZ-Korpus. . . . .	262
13.26	Typische Mehrworteinheiten aus den Ressorts ‚Feuilleton‘, ‚Sport‘ und ‚Wirtschaft‘ in den Zeitperioden 1995–1997 und 2003–2005. . . . .	268

13.27	Prozentuale Anteile des Musters [DEKADE] JAHRE an allen Artikeln des jeweiligen Ressorts (‚Feuilleton‘, ‚Ausland‘ und ‚Inland‘) im NZZ-Korpus 1995–2005. . . . .	272
13.28	Typische Mehrworteinheiten aus dem Ressort ‚Feuilleton‘ in den Zeitperioden 1995–1997 (links) und 2003–2005 (rechts) mit dem Lemma <i>deutsch</i> als Bestandteil. . .	281
13.29	Vorkommen (Einheit: Artikel) von X des Musters [DEF. ART.] DEUTSCHE(N) X im Ressort ‚Feuilleton‘ im Vergleich zwischen den Zeitperioden 1995–1998, 1999–2001 und 2002–2005. . . . .	282
13.30	Vorkommen (Einheit: Artikel) verschiedener Varianten von LESER/IN im Singular und Plural im NZZ-Korpus über die Zeitperiode 1995–2005. . . . .	284
13.31	Vorkommen (Einheit: Artikel) von <i>Damen</i> bzw. <i>Frauen</i> im Ressort ‚Sport‘ und in allen anderen Ressorts im NZZ-Korpus. . . . .	289
13.32	Verteilung und $\chi^2$ -Statistik (Einheit: Artikel) von <i>Es ist</i> über die Ressorts des NZZ-Korpus. . . . .	292
17.1	Typische Mehrworteinheiten in Beiträgen der Autoren A und B im Politikforum.de-Korpus. . . . .	326
17.2	Auswahl von Belegen für das Muster X PROZENT in den Beiträgen der Autoren A und B im Politikforum.de-Korpus. . . . .	328
A.1	Typikprofil Ressort ‚Ausland‘ 1995–1997. . . . .	342
A.2	Typikprofil Ressort ‚Ausland‘ 2003–2005. . . . .	344
A.3	Typikprofil Ressort ‚Inland‘ 1995–1997. . . . .	346
A.4	Typikprofil Ressort ‚Inland‘ 2003–2005. . . . .	349
A.5	Typikprofil Ressort ‚Feuilleton‘ 1995–1997. . . . .	350
A.6	Typikprofil Ressort ‚Feuilleton‘ 2003–2005. . . . .	351
A.7	Typikprofil Ressort ‚Wirtschaft‘ 1995–1997. . . . .	354
A.8	Typikprofil Ressort ‚Wirtschaft‘ 2003–2005. . . . .	356
A.9	Typikprofil Ressort ‚Sport‘ 1995–1997. . . . .	357
A.10	Typikprofil Ressort ‚Sport‘ 2003–2005. . . . .	358
A.11	Typische Mehrworteinheiten im Ressort ‚Leserbriefe‘ des NZZ-Korpus in den Jahren 1995–2005 im Vergleich zu den anderen Ressorts (manuell gefiltert, Top 103). . . . .	360

## Abbildungsverzeichnis

1.1	Kollokatoren zu <i>Tisch</i> im DeReKo IDS (o. J.)-Korpus. . .	1
2.1	Muster bestehen aus Ensemblestücken. . . . .	25
2.2	Unterschiedlich starke Ausprägungen von Mustern. . . .	26
4.1	Das einfache Toulminschema. . . . .	68
4.2	Die Typologie der Topoi der Alltagslogik. . . . .	73
4.3	Einordnung von Konzepten zu Sprachgebrauch in die Kategorien ‚Theorien‘, ‚Phänomene‘ und ‚Ziele‘. . . . .	88
4.4	Spinnenprofile semantischer Kategorien und linguistischer Analysekonzepte . . . . .	90
4.5	Spinnenprofil der Stärken korpuslinguistischer Ansätze. . .	91
5.1	Die Operationalisierung einer korpuslinguistischen Diskursanalyse im Überblick. . . . .	104
8.1	Perspektiven, Methoden und Hintergründe der Interpretation. . . . .	154
8.2	Die relativen Frequenzen in Prozent von Artikeln im NZZ-Korpus, die den Ausdruck <i>Kampf gegen</i> enthalten. . .	156
8.3	Die relativen Frequenzen in Prozent von Artikeln im NZZ-Korpus, die den Ausdruck <i>Kampf gegen (den) Terror</i> (inkl. Flexionsformen) enthalten. . . . .	158
8.4	Die Differenzkoeffizienten D für die Ausdrücke <i>Kampf gegen</i> und <i>Kampf gegen (den) Terror</i> (inkl. Flexionsformen) im NZZ-Korpus (Basis Artikel). . . . .	159
8.5	Die Füllungen von X im Muster KAMPF GEGEN [ARTIKEL] X für die Zeit vor und nach dem 11. September 2001. . .	162
8.6	Paradigmatische Konstellationen von STERBEN im Ressort ‚Ausland‘. . . . .	165
8.7	Paradigmatische Konstellationen von STERBEN über alle Ressorts. . . . .	166

8.8	Signifikante Kollokatoren zu <i>riß</i> bzw. <i>riss</i> im Wortschatz Leipzig (o.J.)-Korpus, berechnet am 21. April 2007. . . .	173
11.1	Umfang der Ressorts pro Jahr im NZZ-Korpus in Anzahl Wörter. . . . .	196
13.1	Paradigmatische Konstellation für ETHNIENBEZEICHNUNGEN und LÄNDERBEZEICHNUNGEN in Relation zu Artikeln mit dem Kontext ‚Kroatien/Bosnien/Serbien‘. . . . .	215
13.2	Paradigmatische Konstellation für ETHNIENBEZEICHNUNGEN und LÄNDERBEZEICHNUNGEN in Relation zu Artikeln mit dem Kontext ‚Kroatien/Bosnien/Serbien‘. . . . .	216
13.3	Die Füllungen von X im Muster KAMPF GEGEN [ARTIKEL] X für die Zeit vor und nach dem 11. September 2001. . .	222
13.4	Die Frequenzen der Artikel im Ressort ‚Ausland‘, die mindestens ein Vorkommen von [Zahl] Mann bzw. [Zahl] Soldaten aufweisen. . . . .	239
13.5	Die Frequenzen der Artikel im Ressort ‚Inland‘ und in allen Ressorts, die mindestens ein Vorkommen von <i>vor den Medien</i> aufweisen. . . . .	243
13.6	Die Frequenzen der Artikel in der NZZ von 1995–2007 (vollständige Daten), die mindestens ein Vorkommen von <i>Klimawandel</i> bzw. <i>Klimaerwärmung</i> aufweisen. . . . .	267
13.7	Ressort ‚Ausland‘: Nennung von Staaten, Hauptstädten und Einwohner/innen im NZZ-Korpus. . . . .	276
13.8	Ressort ‚Wirtschaft‘: Nennung von Staaten, Hauptstädten und Einwohner/innen im NZZ-Korpus. . . . .	276
13.9	Ressort ‚Feuilleton‘: Nennung von Staaten, Hauptstädten und Einwohner/innen im NZZ-Korpus. . . . .	277
13.10	Ressort ‚Ausland‘: Die größten Veränderungen der Nennungen von Staaten, Hauptstädten und Einwohner/innen im NZZ-Korpus. . . . .	278
13.11	Ressort ‚Wirtschaft‘: Die größten Veränderungen der Nennungen von Staaten, Hauptstädten und Einwohner/innen im NZZ-Korpus. . . . .	278
13.12	Ressort ‚Ausland‘: Die größten Veränderungen der Nennungen von Einwohner/innen (Nationalitäten) im NZZ-Korpus. . . . .	279

13.13	Die Frequenzen der Artikel in den Ressorts ‚Sport‘ bzw. in allen anderen Ressorts im NZZ-Korpus, die mindestens ein Vorkommen des Wortteils <i>Swiss</i> aufweisen. . . .	287
13.14	Die Frequenzen der Artikel im Ressort ‚Sport‘ im NZZ-Korpus, die mindestens ein Vorkommen von <i>Damen</i> bzw. <i>Frauen</i> aufweisen. . . . .	288
13.15	Die Frequenzen in den Ressorts des NZZ-Korpus von Artikeln die mindestens ein Vorkommen von <i>Es ist</i> aufweisen. . . . .	292
17.1	Alle genannten Prozentzahlen im Muster X PROZENT in den Beiträgen der Autoren A und B im Politikforum.de-Korpus. . . . .	329



## Literatur

- AG für die Neue Zürcher Zeitung (2005): Geschäftsbericht 2004. Zürich, <http://verlag.nzz.ch/ger/unternehmen/kennzahlen.html> [29. 2. 2008].
- (2006): Geschäftsbericht 2005. Zürich, <http://verlag.nzz.ch/ger/unternehmen/kennzahlen.html> [29. 2. 2008].
- Albert, Ruth/Koster, Cor J. (2002): *Empirie in Linguistik und Sprachlehrforschung. Ein methodologisches Arbeitsbuch*. Tübingen: Narr (Narr Studienbücher).
- Atkins, Sue/Clear, Jeremy/Ostler, Nicholas (1992): Corpus Design Criteria. In: *Literary and Linguistic Computing* 7, H. 1, S. 1–16.
- Baayen, R. Harald (2008): *Analyzing linguistic data: a practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Banerjee, Satanjeev/Pedersen, Ted (2003): The Design, Implementation, and Use of the Ngram Statistic Package. In: *Proceedings of the Fourth International Conference on Intelligent Text Processing and Computational Linguistics*, Mexico City.
- Baroni, Marco/Bernardini, Silvia (Hgg.) (2006): *Wacky! Working papers on the Web as Corpus*. Bologna: GEDIT.
- Belica, Cyril (1996): Analysis of Temporal Changes in Corpora. In: *International Journal of Corpus Linguistics* 1, H. 1, S. 61–73.
- (2001–2006): *Koorkkurrenzdatenbank CCDB. Eine korpuslinguistische Denk- und Experimentierplattform für die Erforschung und theoretische Begründung von systemisch-strukturellen Eigenschaften von Kohäsionsrelationen zwischen den Konstituenten des Sprachgebrauchs*. <http://corpora.ids-mannheim.de/ccdb/> [29. 2. 2008].
- Belica, Cyril/Steyer, Kathrin (2006): Korpusanalytische Zugänge zu sprachlichem Usus. In: *AUC (Acta Universitatis Carolinae), Germanistica Pragensia* XX.
- Berry-Rogghe, Godelive L. M. (1972): Computation of Collocations and their Relevance in Lexical Studies. In: *The Computer and Literary Studies*, hg. v. Angus J. Aitken, R. W. Bailey u. N. Hamilton-Smith, Bd. 103–112.
- Biber, Douglas (1993): Representativeness in Corpus Design. In: *Lit Linguist Computing* 8, H. 4, S. 243–257.
- Biber, Douglas (1994): Representativeness in Corpus Design. In: *Current Issues in Computational Linguistics: in Honour of Don Walker*, S. 377–407 (Linguistica Computazionale; 9–10).
- Bickel, Hans (2006): Das Internet als linguistisches Korpus. In: *Linguistik online* 28, H. 3, S. 71–83, [http://www.linguistik-online.de/28\\_06/bickel.pdf](http://www.linguistik-online.de/28_06/bickel.pdf) [29. 2. 2008].
- Bluhm, Claudia/Deissler, Dirk/Scharloth, Joachim/Stukenbrock, Anja (2000): Linguistische Diskursanalyse: Überblick, Probleme, Perspektiven. In: *Sprache und Literatur in Wissenschaft und Unterricht* 88, S. 3–19.



- Börsenverein des Deutschen Buchhandels (1995): Buch und Buchhandel in Zahlen. Ausgabe 1995. Frankfurt am Main.
- Börsenverein des Deutschen Buchhandels (2000): Buch und Buchhandel in Zahlen. Ausgabe 2000. Frankfurt am Main.
- Börsenverein des Deutschen Buchhandels (2006): Buch und Buchhandel in Zahlen. Ausgabe 2006. Frankfurt am Main.
- Breyer, Yvonne (2005): *Gateway to Corpus Linguistics on the Internet*. Elektronische Ressource, <http://www.corpus-linguistics.de/> [29. 2. 2008].
- Bubenhof, Noah (2006): *Einführung in die Korpuslinguistik: Praktische Grundlagen und Werkzeuge*. Elektronische Ressource, <http://www.bubenhof.com/korpuslinguistik/> [29. 2. 2008].
- (2008a): Diskurse berechnen? Wege zu einer korpuslinguistischen Diskursanalyse. In: *Methoden der Diskurslinguistik. Sprachwissenschaftliche Zugänge zur transtextuellen Ebene*, hg. v. Jürgen Spitzmüller u. Ingo Warnke, Berlin/New York: de Gruyter (Linguistik – Impulse & Tendenzen; 31).
  - (2008b): „Es liegt in der Natur der Sache...“. Korpuslinguistische Untersuchungen zu Kollokationen in Argumentationsfiguren. In: *Studien zur Phraseologie aus textueller Sicht*, hg. v. Carmen Mellado Blanco, Hamburg: Kovač, S. 53–72 (Philologia – Sprachwissenschaftliche Forschungsergebnisse; 112).
- Bubenhof, Noah/Klimke, Martin/Scharloth, Joachim (2008a): *political tracker – U.S. Presidential Campaign '08: A Semantic Matrix Analysis*. Elektronische Ressource, <http://semtracks.com/politicaltracker/>.
- (2008b): *The Word War: „Yes, He Did“. How Obama won the (rhetorical) battle for the White House*. International Relations and Security Network, ISN ETH Zurich, <http://www.isn.ethz.ch/Current-Affairs/Special-Reports/The-Word-War-Yes-He-Did/Analysis>.
  - (2009): *political tracker – Bundestagswahl '09. Eine Semantische Matrixanalyse*. Elektronische Ressource, <http://semtracks.com/politicaltracker/>.
- Bubenhof, Noah/Scharloth, Joachim (2009): Kontext korpuslinguistisch: Die induktive Berechnung von Sprachgebrauchsmustern in großen Textkorpora. In: *Text-Zeichen und Kon-Texte. Studien zu soziokulturellen Konstellationen literalen Handelns*, hg. v. Peter Klotz, Paul R. Portmann-Tselikas u. Georg Weidacher, Tübingen [im Druck].
- Burger, Harald (1998): *Phraseologie. Eine Einführung am Beispiel des Deutschen*. Berlin: Erich Schmidt (Grundlagen der Germanistik; 36).
- Burkhardt, Armin (2003): *Das Parlament und seine Sprache. Studien zu Theorie und Geschichte parlamentarischer Kommunikation*. Tübingen: Niemeyer (Reihe Germanistische Linguistik; 241).
- Busse, Dietrich (1997): Das Eigene und das Fremde. Annotationen zu Funktion und Wirkung einer diskurssemantischen Grundfigur. In: *Die Sprache des Migrationsdiskurses. Das Reden über „Ausländer“ in Medien, Politik und Alltag*, hg. v. Matthias Jung, Martin Wengeler u. Karin Böke, Opladen: Westdeutscher Verlag, S. 17–35.
- (2005): Sprachwissenschaft als Sozialwissenschaft? In: *Brisante Semantik. Neuere Konzepte und Forschungsergebnisse einer kulturwissenschaftlichen Linguistik*, hg. v. Dietrich Busse, Thomas Niehr u. Martin Wengeler, Tübingen: Niemeyer, S. 21–43 (Reihe Germanistische Linguistik; 259).

- Busse, Dietrich/Hermanns, Fritz/Teubert, Wolfgang (Hgg.) (1994): *Begriffsgeschichte und Diskursgeschichte. Methodenfragen und Forschungsergebnisse der historischen Semantik*. Opladen: Westdeutscher Verlag.
- Busse, Dietrich/Teubert, Wolfgang (1994): Ist Diskurs ein sprachwissenschaftliches Objekt? Zur Methodenfrage der historischen Semantik. In: *Begriffsgeschichte und Diskursgeschichte. Methodenfragen und Forschungsergebnisse der historischen Semantik*, hg. v. Dietrich Busse, Fritz Hermanns u. Wolfgang Teubert, Opladen: Westdeutscher Verlag, S. 10–28.
- Carstensen, Kai-Uwe/Ebert, Christian/Endriss, Cornelia/Jekat, Susanne/Klabunde, Ralf/Langer, Hagen (Hgg.) (2001): *Computerlinguistik und Sprachtechnologie. Eine Einführung*. Heidelberg/Berlin: Spektrum.
- Cheng, Winnie/Greaves, Chris/Warren, Martin (2006): From N-Gram to Skip-gram to Concgram. In: *International Journal of Corpus Linguistics* 11, S. 411–433(23).
- Cherubim, Dieter (1980): Zum Programm einer historischen Sprachpragmatik. In: *Ansätze zu einer pragmatischen Sprachgeschichte. Zürcher Kolloquium 1978*, hg. v. Horst Sitta, Tübingen: Niemeyer, S. 3–22 (Reihe Germanistische Linguistik; 21).
- Christ, Oli/Schulze, B. M. (1995): Ein flexibles und modulares Anfragesystem für Textcorpora. In: *Tagungsbericht des Arbeitstreffen Lexikon + Text*, Tübingen: Niemeyer.
- Church, Kenneth W./Gale, William A. (1995): Poisson Mixtures. In: *Natural Language Engineering* 1, H. 2, S. 163–190.
- COSMAS II (2007): *COSMAS II: Die zweite Generation des Korpusrecherche- und -analysesystems COSMAS (Corpus Search, Management and Analysis System)*. Elektronische Ressource, <http://www.ids-mannheim.de/cosmas2/> [29. 2. 2008].
- Duden (1999): *Duden. Das große Wörterbuch der deutschen Sprache in zehn Bänden*. Mannheim, Leipzig, Wien, Zürich: Dudenverlag, 3., völlig neu bearb. und erw. Aufl.
- Duffner, Rolf/Näf, Anton (2006): Digitale Textdatenbanken im Vergleich. In: *Linguistik online* 28, H. 3, S. 7–22, [http://www.linguistik-online.net/28\\_06/duffnerNaef.pdf](http://www.linguistik-online.net/28_06/duffnerNaef.pdf) [29. 2. 2008].
- Dunning, Ted E. (1993): Accurate Methods for the Statistics of Surprise and Coincidence. In: *Computational Linguistics* 19, H. 1, S. 61–74.
- Eco, Umberto (1977): *Zeichen. Einführung in einen Begriff und seine Geschichte*. Frankfurt am Main: Suhrkamp (es; 895).
- (2002): *Einführung in die Semiotik*. München: Fink, 9., unveränderte Aufl.
- Eggler, Marcel (2006): *Argumentationsanalyse textlinguistisch. Argumentative Figuren für und wider den Golfkrieg von 1991*. Tübingen: Niemeyer (Reihe Germanistische Linguistik; 268).
- Evert, Stefan (2005): *The Statistics of Word Cooccurrences. Word Pairs and Collocations*. Phil. Diss. Institut für maschinelle Sprachverarbeitung, Universität Stuttgart.
- Feilke, Helmuth (1993): Sprachlicher Common sense und Kommunikation. Über den ‚gesunden Menschenverstand‘, die Prägung der Kompetenz und die idiomatische Ordnung des Verstehens. In: *Der Deutschunterricht* VI, S. 6–21.

- (1994): *Common sense-Kompetenz. Überlegungen zu einer Theorie ‚sympathischen‘ und ‚natürlichen‘ Meinens und Verstehens*. Frankfurt am Main: Suhrkamp.
- (1996): *Sprache als soziale Gestalt. Ausdruck, Prägung und die Ordnung der sprachlichen Typik*. Frankfurt am Main: Suhrkamp.
- (2000): Die pragmatische Wende in der Textlinguistik. In: *Text- und Gesprächslinguistik/Linguistics of Text and Conversation*, hg. v. Klaus Brinker, Berlin/New York: de Gruyter, Bd. 1, S. 64–82 (Handbücher zur Sprach- und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science; 16).
- (2003): Textroutine, Textsemantik und sprachliches Wissen. In: *Sprache und mehr. Ansichten einer Linguistik der sprachlichen Praxis*, hg. v. Angelika Linke, Hanspeter Ortner u. Paul R. Portmann-Tselikas, Tübingen: Niemeyer, S. 209–230 (Reihe Germanistische Linguistik; 245).
- Fiehler, Reinhard/Barden, Birgit/Elstermann, Mechtild/Kraft, Barbara (2004): *Eigenschaften gesprochener Sprache*. Tübingen: Narr.
- Fillmore, Charles (1976): Pragmatics and the Description of Discourse. In: *Pragmatik/Pragmatics 2. Zur Grundlegung einer expliziten Pragmatik*, hg. v. Siegfried J. Schmidt, München: Fink, S. 83–104 (Kritische Information; 25).
- Firth, John Rupert (1957): Modes of Meaning. In: *Papers in Linguistics 1934–1951*, London: Oxford University Press, S. 190–215.
- Fix, Ulla (1991): Stilistische Textanalyse – immer ein Vergleich? Das Gemeinsame von Methoden der Stilanalyse – das Gemeinsame an Stilbegriffen. In: *Aspekte der Textlinguistik*, hg. v. Klaus Brinker, Hildesheim, Zürich, New York: Georg Olms, S. 133–156.
- (1996): Gestalt und Gestalten. Von der Notwendigkeit der Gestaltkategorie für eine das Ästhetische berücksichtigende pragmatische Stilistik. In: *Zeitschrift für Germanistik* 2, S. 308–323.
- Fletcher, William H. (2007): Implementing a BNC-Compare-able Web Corpus. In: *Building and Exploring Web Corpora – Proceedings of the 3rd Web as Corpus Workshop, Incorporating CleanEval (WAC3-2007, September 2007)*, UCL, hg. v. C. Fairon, H. Naets, A. Kilgariff u. G-M de Schrijver, Louvain: Presses Universitaires de Louvain.
- Foucault, Michel (1981): *Archäologie des Wissens*. Frankfurt am Main: Suhrkamp, 10. Aufl.
- (2000): *Die Ordnung des Diskurses*. Frankfurt am Main: Fischer.
- Geertz, Clifford (1987a): Common sense als kulturelles System. In: *Dichte Beschreibung. Beiträge zum Verstehen kultureller Systeme*, Frankfurt am Main: Suhrkamp, S. 261–288 (stw; 696).
- (1987b): Dichte Beschreibung. Bemerkungen zu einer deutenden Theorie von Kultur. In: *Dichte Beschreibung. Beiträge zum Verstehen kultureller Systeme*, Frankfurt am Main: Suhrkamp, S. 7–43 (stw; 696).
- Geideck, Susan/Liebert, Wolf-Andreas (2003): Sinnformeln. Eine soziologisch-linguistische Skizze. In: *Sinnformeln. Linguistische und soziologische Analysen von Leitbildern, Metaphern und anderen kollektiven Orientierungsmustern*, hg. v. Susan Geideck u. Wolf-Andreas Liebert, Berlin, New York: de Gruyter, S. 3–14 (Linguistik – Impulse & Tendenzen; 2).

- von Glaserfeld, Ernst (1992): Konstruktion der Wirklichkeit und des Begriffs der Objektivität. In: *Einführung in den Konstruktivismus*, hg. v. Heinz Gumin u. Heinrich Meier, München, Zürich: Piper, S. 9–39.
- Glasze, Georg (2007): Vorschläge zur Operationalisierung der Diskurstheorie von Laclau und Mouffe in einer Triangulation von lexikometrischen und interpretativen Methoden. In: *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 8, H. 2.
- Glasze, Georg/Mattisek, Annika (2009): *Handbuch Diskurs und Raum: Theorien und Methoden für die Humangeographie sowie die sozial- und kulturwissenschaftliche Raumforschung*. Bielefeld: Transcript Verlag.
- Günthner, Susanne (1995): Gattungen in der sozialen Praxis. In: *Deutsche Sprache* 3, S. 193–218.
- (2003): Eine Sprachwissenschaft der „lebendigen Rede“. Ansätze einer Anthropologischen Linguistik. In: *Sprache und mehr. Ansichten einer Linguistik der sprachlichen Praxis*, hg. v. Angelika Linke, Hanspeter Ortner u. Paul R. Portmann-Tselikas, Tübingen: Niemeyer, S. 189–208 (Reihe Germanistische Linguistik; 245).
- Günthner, Susanne/Knoblach, Hubert (1994): „Forms are the Food of Faith“. Gattungen als Muster kommunikativen Handelns. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 46, S. 693–723.
- (1995): Culturally Patterned Speaking Practices. The Analysis of Communicative Genres. In: *Pragmatics* 5, H. 1, S. 1–32.
- Halliday, Michael Alexander Kirkwood (1961): Categories of the Theory of Grammar. In: *Word* 17, S. 241–292.
- Halliday, Michael Alexander Kirkwood/Teubert, Wolfgang/Yallop, Colin/Čermáková, Anna (2004): *Lexicology and Corpus Linguistics. An Introduction*. London/New York: Continuum.
- Haugen, Einar (1958): (Review of) Papers in Linguistics 1934–1951. By J. R. Firth. In: *Language* 34, H. 4, S. 498–502.
- Hausmann, Franz Josef (1985): Kollokationen im deutschen Wörterbuch. Ein Beitrag zur Theorie des lexikographischen Beispiels. In: *Lexikographie und Grammatik. Akten des Essener Kolloquiums zur Grammatik im Wörterbuch 1984*, hg. v. H. Bergenholtz u. J. Mugdan, Tübingen, S. 118–129 (Lexicographica Series Maior; 3).
- Herberg, Dieter (1998): Neues im Wortgebrauch der Wendezeit. Zur Arbeit mit dem IDS-Wendekorpus. In: *Neologie und Korpus*, hg. v. Wolfgang Teubert, Tübingen: Narr, S. 43–62 (Studien zur deutschen Sprache; 11).
- Hermanns, Fritz (1995): Sprachgeschichte als Mentalitätsgeschichte. Überlegungen zu Sinn und Form und Gegenstand historischer Semantik. In: *Sprachgeschichte des Neuhochdeutschen. Gegenstände, Methoden, Theorien*, hg. v. Andreas Gardt, Klaus Mattheier u. Oskar Reichmann, Tübingen: Niemeyer, S. 69–101.
- Hinnenkamp, Volker/Selting, Margret (Hgg.) (1989): *Stil und Stilisierung. Arbeiten zur interpretativen Soziolinguistik*. Tübingen: Niemeyer.
- Humboldt, Wilhelm von (1973): *Schriften zur Sprache*. Stuttgart: Reclam.
- Institut für Deutsche Sprache (2004): *Eine kurze Einführung in die Kookkurrenzanalyse und syntagmatische Muster*. Elektronische Ressource, <http://www.ids-mannheim.de/kl/misc/tutorial.html> [29. 2. 2008].

- Jäger, Ludwig (2003): Erkenntnisobjekt Sprache. Probleme der linguistischen Gegenstandskonstitution. In: *Sprache und mehr. Ansichten einer Linguistik der sprachlichen Praxis*, hg. v. Angelika Linke, Hanspeter Ortner u. Paul R. Portmann-Tselikas, Tübingen: Niemeyer, S. 67–98 (Reihe Germanistische Linguistik; 245).
- (2005): Vom Eigensinn des Mediums Sprache. In: *Brisante Semantik. Neuere Konzepte und Forschungsergebnisse einer kulturwissenschaftlichen Linguistik*, hg. v. Dietrich Busse, Thomas Niehr u. Martin Wengeler, Tübingen: Niemeyer, S. 45–64 (Reihe Germanistische Linguistik; 259).
- Jones, Seth G./Libicki, Martin C (2008): *How terrorist groups end: lessons for countering Al Qa'ida*. Santa Monica, CA: Rand.
- Jung, Matthias (1996): Linguistische Diskursgeschichte. In: *Öffentlicher Sprachgebrauch. Praktische, theoretische und historische Perspektiven. Georg Stötzel zum 60. Geburtstag gewidmet*, hg. v. Karin Böke, Matthias Jung u. Martin Wengeler, Opladen: Westdeutscher Verlag, S. 453–472.
- (2005): Schlüsselwortforschung im Internet – Möglichkeiten, Beispiele, Grenzen. In: *Brisante Semantik. Neuere Konzepte und Forschungsergebnisse einer kulturwissenschaftlichen Linguistik*, hg. v. Dietrich Busse, Thomas Niehr u. Martin Wengeler, Tübingen: Niemeyer, S. 355–368 (Reihe Germanistische Linguistik; 259).
- Keibel, Holger/Belica, Cyril (2007): CCDB: A Corpus-Linguistic Research and Development Workbench. In: *Proceedings of the 4th Corpus Linguistics Conference*, Birmingham.
- Kienpointner, Manfred (1992): *Alltagslogik. Struktur und Funktion von Argumentationsmustern*. Stuttgart-Bad Cannstatt: Frommann-Holzboog (Problemmata; 126).
- Kilgarriff, Adam (1996a): Which Words are Particularly Characteristic of a Text? A Survey of Statistical Approaches. In: *Proc. AISB Workshop on Language Engineering for Document Analysis and Recognition, Sussex University, April 1996*, Sussex University.
- (1996b): Why Chi-Square doesn't Work, and an Improved LOB-Brown Comparison. In: *Proc. ALLC-ACH Conference*, Bergen, S. 169–172.
  - (2001): Comparing Corpora. In: *International Journal of Corpus Linguistics* 6, H. 1, S. 1–37.
  - (2005): Language is Never, Ever, Ever, Random. In: *Corpus Linguistics and Linguistic Theory* 1, H. 2, S. 263–276.
- Kilgarriff, Adam/Grefenstette, Gregory (2003): Introduction to the Special Issue on the Web as Corpus. In: *Computational Linguistics* 29, H. 3, S. 333–347.
- Kilgarriff, Adam/Rose, Tony (1998): Measures for Corpus Similarity and Homogeneity. In: *Proc. 3rd Conf. on Empirical Methods in Natural Language Processing (EMNLP-3). Granada, Spain, June 1998*, S. 46–52.
- Kneer, Georg/Nassehi, Armin (1994): *Niklas Luhmanns Theorie sozialer Systeme: eine Einführung*. Bd. 1751. München: Fink, 2., unveränd. Aufl. (Uni-Taschenbücher. Soziologie, Philosophie, Literaturwissenschaft).
- Koller, Veronika (2006): Of Critical Importance: Using Electronic Text Corpora to Study Metaphor in Business Media Discourse. In: *Corpus-Based Approaches to Metaphor and Metonymy*, hg. v. Anatol Stefanowitsch u. Stefan Th. Gries, Berlin: de Gruyter, S. 237–266 (Trends in linguistics. Studies and mono-

- graphs; 171).
- Kopperschmidt, Josef (1989): *Methodik der Argumentationsanalyse*. Stuttgart-Bad Cannstatt: Frommann-Holzboog (Problemata; 119).
- (1991): Formale Topik. Anmerkungen zu ihrer heuristischen Funktionalisierung innerhalb einer Argumentationsanalytik. In: *Rhetorik zwischen den Wissenschaften. Geschichte, System, Praxis als Probleme des „Historischen Wörterbuchs der Rhetorik“*, hg. v. Gert Ueding, Tübingen: Niemeyer, S. 53–62 (Rhetorik-Forschung; 1).
  - (2000): *Argumentationstheorie zur Einführung*. Hamburg: Junius.
- Kövecses, Zoltan (1998): Are there any Emotion-Specific Metaphors? In: *Speaking of Emotions. Conceptualization and Expression*, hg. v. Angeliki Athanasiadou u. Elzbieta Tabaskowska, Berlin, New York: de Gruyter, S. 127–151.
- Labov, William (1972): *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Lakoff, George (1991): Metapher und Krieg. In: *Sprache im technischen Zeitalter* 119, S. 221–239.
- Lakoff, George/Johnson, Mark (1980): *Metaphors we Live by*. Chicago: University of Chicago Press.
- (1998): *Leben in Metaphern. Konstruktion und Gebrauch von Sprachbildern*. Heidelberg: Carl-Auer-Systeme.
- Lehr, Andrea (1996): *Kollokationen und maschinenlesbare Korpora. Ein operationales Analysemodell zum Aufbau lexikalischer Netze*. Tübingen: Niemeyer (Reihe Germanistische Linguistik; 168).
- Lemmitzer, Lothar (1997): *Extraktion komplexer Lexeme aus Textkorpora*. Tübingen: Niemeyer (Reihe Germanistische Linguistik; 180).
- Lemmitzer, Lothar/Zinsmeister, Heike (2006): *Korpuslinguistik. Eine Einführung*. Tübingen: Narr.
- Linke, Angelika (1996): *Sprachkultur und Bürgertum: zur Mentalitätsgeschichte des 19. Jahrhunderts*. Stuttgart: Metzler.
- (2001): Trauer, Öffentlichkeit und Intimität. Zum Wandel der Textsorte ‚Todesanzeige‘ in der zweiten Hälfte des 20. Jahrhunderts. In: *Zur Kulturspezifik von Textsorten*, hg. v. Ulla Fix, Stephan Habscheid u. Josef Klein, Tübingen: Niemeyer, S. 195–223 (Textsorten; 3).
  - (2002): Senioren – Zur Konstruktion von (Alters-?)Gruppen im Medium Sprache. In: *Spracherwerb und Lebensalter*, hg. v. Annelies Häcki-Buhofer, Tübingen: Francke, S. 21–36.
  - (2003a): Begriffsgeschichte – Diskursgeschichte – Sprachgebrauchsgeschichte. In: *Herausforderungen der Begriffsgeschichte*, hg. v. Carsten Dutt, Heidelberg: Winter, S. 39–49.
  - (2003b): Spaß haben. Ein Zeitgefühl. In: *Standardfragen. Soziolinguistische Perspektiven auf Sprachgeschichte, Sprachkontakt und Sprachvariation*, hg. v. Jannis K. Androutsopoulos u. Evelyn Ziegler, Frankfurt am Main: Lang.
  - (2003c): Sprachgeschichte – Gesellschaftsgeschichte – Kulturanalyse. In: *Germanistische Linguistik: Konturen eines Faches*, hg. v. Helmut Henne, Horst Sitta u. Herbert Ernst Wiegand, Tübingen: Niemeyer.
  - (2004): Das Unbeschreibliche. Zur Sozialesemiotik adeligen Körperverhaltens im 18. und 19. Jahrhundert. In: *Adel und Moderne. Deutschland im europäischen Vergleich im 19. und 20. Jahrhundert*, hg. v. Eckart Conze u. Monika

- Wienfort, Köln, Weimar, Wien: Böhlau, S. 247–268.
- (2006): Sprachliche Amerikanisierung und Popular Culture. Zur kulturellen Deutung fremder Zeichen. In: *Attraktion und Abwehr. Die Amerikanisierung der Alltagskultur in Europa*, hg. v. Angelika Linke u. Jakob Tanner, Köln, Weimar, Wien: Böhlau, S. 37–51.
  - (2008): Kommunikation, Kultur und Vergesellschaftung. Überlegungen zu einer Kulturgeschichte der Kommunikation. In: *Jahrbuch 2007 Institut für Deutsche Sprache, Mannheim*, Berlin, New York: de Gruyter.
  - (2009): ‚Varietät‘ vs. ‚kommunikative Praktik‘ – welcher Zugang nützt der Sprachgeschichte? In: *Empirische Evidenzen und theoretische Passungen sprachlicher Variation*, hg. v. Peter Gilles, Joachim Scharloth u. Evelyn Ziegler, Frankfurt am Main: Lang [im Druck].
- Linke, Angelika/Günthner, Susanne (2006): Einleitung: Linguistik und Kulturanalyse. Ansichten eines symbiotischen Verhältnisses. In: *Zeitschrift für germanistische Linguistik* 34, S. 1–27.
- Longchamp, Claude (2002): Wirtschaftsskepsis bestimmt die heutigen Sorgen. Schlussbericht zum „Sorgenbarometer 2002“ für das Bulletin der CS. Elektronische Ressource, Bern, <http://emagazine.credit-suisse.com> [29. 2. 2008].
- Longchamp, Claude/Golder, Lukas (2003): Das Sorgenbarometer 2003 der Credit Suisse. „Sorge um den Arbeitsplatz dominiert“. Elektronische Ressource, Bern, <http://emagazine.credit-suisse.com> [29. 2. 2008].
- (2004): Sorgenbarometer der Credit Suisse 2004. Arbeitslosigkeit, Gesundheitswesen, Altersvorsorge und Asylwesen. Schlussbericht zum allgemeinen Teil des Sorgenbarometers 2004, im Auftrag vom Bulletin der Credit Suisse. Elektronische Ressource, Bern, <http://emagazine.credit-suisse.com> [29. 2. 2008].
  - (2005): Die Arbeitslosigkeit und ihre Folgen für Problembewusstsein, Wirtschaftsempfinden und Institutionenvertrauen. Schlussbericht zum allgemeinen Teil des Sorgenbarometers 2005, im Auftrag vom Bulletin der Credit Suisse. Elektronische Ressource, Bern, <http://emagazine.credit-suisse.com> [29. 2. 2008].
- Longchamp, Claude/Leuenberger, Petra (1999): Wer nichts riskiert, gewinnt nichts. Eine Analyse des Bewusstseins öffentlicher Probleme in den 90er Jahren. Bericht zum „Sorgenbarometer 1999“ des GfS-Forschungsinstituts. Elektronische Ressource, Bern, <http://emagazine.credit-suisse.com> [29. 2. 2008].
- (2000): Der Trend zum Bewusstwerden sozialpolitischer Probleme. Bericht zum „Sorgenbarometer 2000“ für das Bulletin der CS. Elektronische Ressource, Bern, <http://emagazine.credit-suisse.com> [29. 2. 2008].
  - (2001): Ereignisse, politische und wirtschaftliche Zyklen als Determinanten des Problembewusstseins. Schlussbericht zum „Sorgenbarometer 2001“ für das Bulletin der CS. Elektronische Ressource, Bern, <http://emagazine.credit-suisse.com> [29. 2. 2008].
- Luckmann, Thomas (1986): Grundformen der gesellschaftlichen Vermittlung des Wissens: Kommunikative Gattungen. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* Sonderheft 27, S. 191–211.
- Lyne, Anthony A. (1985): *The Vocabulary of French Business Correspondence*. Geneva, Paris: Slatkine-Champion.

- Manning, Christopher D./Schütze, Hinrich (2002): *Foundations of Statistical Natural Language Processing*. Cambridge, Massachusetts: The MIT Press, 5. Aufl.
- Mautner, Gerlinde (2005): Time to get Wired: Using Web-Based Corpora in Critical Discourse Analysis. In: *Discourse & Society* 6, H. 16.
- Müller, Ralph (2006): Kreative Metaphern über Europa. Eine konzeptuelle Analyse. In: *Der Deutschunterricht* 6, S. 53–61.
- Musolff, Andreas (2005): Brisante Metaphern. Zur argumentativen Funktion von Krankheitsmetaphorik im öffentlichen Diskurs. In: *Brisante Semantik. Neuere Konzepte und Forschungsergebnisse einer kulturwissenschaftlichen Linguistik*, hg. v. Dietrich Busse, Thomas Niehr u. Martin Wengeler, Tübingen: Niemeyer, S. 309–322 (Reihe Germanistische Linguistik; 259).
- Nerlich, Brigitte (1995): The 1930s – At the Birth of a Pragmatic Conception of Language. In: *Historiographica Linguistica* XXII, H. 3, S. 311–334.
- Nünning, Ansgar/Nünning, Vera (Hgg.) (2003): *Konzepte der Kulturwissenschaften*. Stuttgart, Weimar: Metzler.
- NZZ Mediadok (2005): Ergebnisse aus dem Medien-Radar 2005. Zürich, [http://verlag.nzz.ch/pdf/documentations\\_16\\_1.pdf](http://verlag.nzz.ch/pdf/documentations_16_1.pdf) [29. 2. 2008].
- Oakes, Michael (1998): *Statistics for Corpus Linguistics*. Edinburgh: Edinburgh University Press (Edinburgh Textbooks in Empirical Linguistics).
- Partington, Alan (2006): Metaphors, Motifs and Similes across Discourse Types: Corpus-Assisted Discourse Studies (CADS) at work. In: *Corpus-Based Approaches to Metaphor and Metonymy*, hg. v. Anatol Stefanowitsch u. Stefan Th. Gries, Berlin: de Gruyter, S. 267–304 (Trends in linguistics. Studies and monographs; 171).
- Pearce, Darren (2002): A Comparative Evaluation of Collocation Extraction Techniques. In: *Proceedings of the 3rd International Conference on Language Resources and Evaluation, LREC'02*, Las Palmas de Gran Canaria, Spain, S. 1530–1536.
- Pedersen, Ted (1996): Fishing for Exactness. In: *Proceedings of the Conference of South-Central SAS Users Group*, hg. v. SAS Users Group, Texas.
- Perkuhn, Rainer (2007): ‚Corpus-Driven‘: Systematische Auswertung automatisch ermittelter sprachlicher Muster. In: *Sprach-Perspektiven. Germanistische Linguistik und das Institut für Deutsche Sprache*, hg. v. Ludwig M. Eichinger u. Heidrun Kämper, Tübingen: Narr, S. 465–491 (Studien zur deutschen Sprache; 40).
- Perkuhn, Rainer/Belica, Cyril (2006): Korpuslinguistik – Das unbekannte Wesen. Oder Mythen über Korpora und Korpuslinguistik. In: *Sprachreport* 22, H. 1, S. 2–8.
- Perkuhn, Rainer/Belica, Cyril/al Wadi, Doris/Lauer, Meike/Steyer, Kathrin/Weiß, Christian (2005): Korpustechnologie am Institut für Deutsche Sprache. In: *Korpuslinguistik deutsch: synchron – diachron – kontrastiv. Würzburger Kolloquium 2003*, hg. v. Johannes Schwitalla u. Werner Wegstein, Tübingen: Niemeyer, S. 57–70.
- Quasthoff, Uwe/Richter, Matthias/Biemann, Christian (2006): Corpus Portal for Search in Monolingual Corpora. In: *Proceedings of the fifth international conference on Language Resources and Evaluation, LREC 2006, Genoa*, S. 1799–1802.



- Rayson, Paul (2003): *Matrix: A Statistical Method and Software Tool for Linguistic Analysis through Corpus Comparison*. Phil. Diss. Lancaster University, <http://www.comp.lancs.ac.uk/computing/users/paul/public.html#phd> [29. 2. 2008].
- Rayson, Paul/Garside, Roger (2000): Comparing Corpora using Frequency Profiling. In: *Proceedings of the workshop on Comparing corpora*, Morristown, NJ, USA: Association for Computational Linguistics, S. 1–6.
- Rietveld, Toni/Hout, Roeland van (2005): *Statistics in Language Research: Analysis of Variance*. Berlin, New York: de Gruyter.
- Rolf, Eckard (2005): *Metaphertheorien. Typologie, Darstellung, Bibliographie*. Berlin, New York: de Gruyter.
- Sandig, Barbara (1995): Tendenzen der linguistischen Stilforschung. In: *Stilfragen*, hg. v. G. Stickel, Berlin, New York: de Gruyter, S. 27–61.
- Savigny, Eike von (1994): Wittgensteins „*Philosophische Untersuchungen*“. *Ein Kommentar für Leser*. Frankfurt am Main: Klostermann, 2. völlig überarbeitete und vermehrte Aufl. [Zwei Bände: Band 1 Abschnitte 1 bis 315, Band 2 Abschnitte 316 bis 693].
- Scharloth, Joachim (2005a): Die Semantik der Kulturen. Diskurssemantische Grundfiguren als Kategorien einer linguistischen Kulturanalyse. In: *Brisante Semantik. Neuere Konzepte und Forschungsergebnisse einer kulturwissenschaftlichen Linguistik*, hg. v. Dietrich Busse, Thomas Niehr u. Martin Wengeler, Tübingen: Niemeyer, S. 133–148 (Reihe Germanistische Linguistik; 259).
- (2005b): *Sprachnormen und Mentalitäten. Sprachbewusstseinsgeschichte in Deutschland im Zeitraum von 1766 bis 1785*. Tübingen: Niemeyer (Reihe Germanistische Linguistik; 255).
- Scherer, Carmen (2006): *Korpuslinguistik*. Heidelberg: Winter (Kurze Einführungen in die Germanistische Linguistik; 2).
- Schiller, Anne/Teufel, Simone/Thielen, Christine (1995): Guidelines für das Tagging deutscher Textcorpora mit STTS. Universität Stuttgart, Institut für maschinelle Sprachverarbeitung; Universität Tübingen, Seminar für Sprachwissenschaft, Stuttgart, <http://www.sfs.uni-tuebingen.de/Elwis/stts/stts.html>.
- Schmid, Helmut (1994): *Probabilistic Part-of-Speech Tagging Using Decision Trees*. <http://www.ims.uni-stuttgart.de/ftp/pub/corpora/tree-tagger1.pdf> [29. 2. 2008].
- (1995): *Improvements in Part-of-Speech Tagging with an Application to German*. <http://www.ims.uni-stuttgart.de/ftp/pub/corpora/tree-tagger2.pdf> [29. 2. 2008].
- Selting, Margret/Hinnenkamp, Volker (1989): Einleitung: Stil und Stilisierung in der Interpretativen Soziolinguistik. In: *Stil und Stilisierung. Arbeiten zur interpretativen Soziolinguistik*, hg. v. Volker Hinnenkamp u. Margret Selting, Tübingen: Niemeyer, S. 1–26 (Linguistische Arbeiten; 235).
- Sharoff, Serge (2006): Open-source Corpora: Using the Net to Fish for Linguistic Data. In: *International Journal of Corpus Linguistics* 11, S. 435–462(28).
- Sinclair, John (1991): *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- (1992): The Automatic Analysis of Corpora. In: *Directions in Corpus Linguistics*, hg. v. Jan Svartvik, Berlin, New York: de Gruyter, S. 379–397 (Trends in Linguistics; 65).

- (1996): The Search for Units of Meaning. In: *TEXTUS* 9, H. 1, S. 75–106.
- (2004): *Trust the Text. Language, Corpus and Discourse*. London: Routledge.
- Spitzmüller, Jürgen (2005): *Metasprachdiskurse. Einstellungen zu Anglizismen und ihre wissenschaftliche Rezeption*. Berlin: de Gruyter (Linguistik – Impulse & Tendenzen; 11).
- Stefanowitsch, Anatol (2006a): Corpus-Based Approaches to Metaphor and Metonymy. In: *Corpus-Based Approaches to Metaphor and Metonymy*, hg. v. Anatol Stefanowitsch u. Stefan Th. Gries, Berlin: de Gruyter, S. 1–16 (Trends in Linguistics. Studies and Monographs; 171).
- (2006b): Words and their Metaphors: A Corpus-Based Approach. In: *Corpus-Based Approaches to Metaphor and Metonymy*, hg. v. Anatol Stefanowitsch u. Stefan Th. Gries, Berlin: de Gruyter, S. 63–105 (Trends in Linguistics. Studies and Monographs; 171).
- Stefanowitsch, Anatol/Gries, Stefan Th. (2003): Collostructions: Investigating the Interaction of Words and Constructions. In: *International Journal of Corpus Linguistics* 8, H. 2, S. 209–243.
- Stefanowitsch, Anatol/Gries, Stefan Thomas (2006): *Corpus-Based Approaches to Metaphor and Metonymy*. Berlin: de Gruyter (Trends in Linguistics. Studies and Monographs; 171).
- Steyer, Kathrin (2004a): Kookkurrenz. Korpusmethodik, linguistisches Modell, lexikografische Perspektiven. In: *Wortverbindungen – mehr oder weniger fest*, hg. v. Kathrin Steyer, Berlin, New York: de Gruyter, S. 87–116 (Institut für Deutsche Sprache. Jahrbuch 2003).
- (Hg.) (2004b): *Wortverbindungen – mehr oder weniger fest*. Berlin/New York: de Gruyter.
- Steyer, Kathrin/Brunner, Annelen (2009): *Das UWV-Analysemodell. Eine korpusgesteuerte Methode zur linguistischen Systematisierung von Wortverbindungen*. Institut für Deutsche Sprache (Online publizierte Arbeiten zur Linguistik OPAL; 1).
- Steyer, Kathrin/Lauer, Meike (2007): „Corpus-Driven“: Linguistische Interpretation von Kookkurrenzbeziehungen. In: *Sprach-Perspektiven. Germanistische Linguistik und das Institut für Deutsche Sprache*, hg. v. Ludwig M. Eichinger u. Heidrun Kämper, Tübingen: Narr, S. 493–509 (Studien zur deutschen Sprache; 40).
- Stocker, Christa (2005): *Sprachgeprägte Frauenbilder. Soziale Stereotype im Mädchenbuch des 19. Jahrhunderts und ihre diskursive Konstituierung*. Tübingen: Niemeyer (Reihe Germanistische Linguistik; 262).
- Stukenbrock, Anja (2005): *Sprachnationalismus. Sprachreflexion als Medium kollektiver Identitätsstiftung in Deutschland (1617–1945)*. Berlin, New York: de Gruyter (Studia Linguistica Germanica; 74).
- Teubert, Wolfgang (2005): My Version of Corpus Linguistics. In: *International Journal of Corpus Linguistics* 10, H. 1, S. 1–13.
- Thelwall, Mike (2005): Creating and Using Web Corpora. In: *International Journal of Corpus Linguistics* 10, H. 4, S. 517–541.
- Tognini-Bonelli, Elena (2001): *Corpus Linguistics at Work*. Amsterdam: Benjamins (Studies in Corpus linguistics; 6).
- Toulmin, Stephen (1958): *The Uses of Argument*. Cambridge: Cambridge University Press.

- Toutenburg, Helge (2000a): *Deskriptive Statistik. Eine Einführung mit SPSS für Windows mit Übungsaufgaben und Lösungen*. Berlin, New York: Springer.
- (2000b): *Induktive Statistik. Eine Einführung mit SPSS für Windows*. Berlin, New York: Springer, 2. neu bearb. und erw. Aufl.
- Tuldava, Juhan (2005): Stylistics, Author Identification. In: *Quantitative Linguistik/Quantitative Linguistics*, hg. v. Reinhard Köhler, Gabriel Altmann u. Rajmund G. Piotrowski, Berlin, New York: de Gruyter, S. 368–387 (Handbücher zur Sprach- und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science; 27).
- Tummers, Jose/Heylen, Kris/Geeraerts, Dirk (2005): Usage-Based Approaches in Cognitive Linguistics: A Technical State of the Art. In: *Corpus Linguistics and Linguistic Theory* 1, H. 2, S. 225–261.
- Volk, Martin (2001): The Automatic Resolution of Prepositional Phrase-Attachment Ambiguities in German. Zürich [Habilitation Thesis].
- Volk, Martin (2002): Using the Web as Corpus for Linguistic Research. In: *Täendusepüüdja. Catcher of the Meaning. A Festschrift for Professor Haldur Õim*, hg. v. Tiit Hennoste u. Renate Pajusalu, University of Tartu (Publications of the Department of General Linguistics; 3).
- Warnke, Ingo (2001): Intrakulturell vs. interkulturell – Zur kulturellen Bedingtheit von Textmustern. In: *Zur Kulturspezifik von Textsorten*, hg. v. Ulla Fix, Stephan Habscheid u. Josef Klein, Tübingen: Stauffenburg, S. 241–254 (Textsorten; 3).
- (2002): Adieu Text – bienvenue Diskurs? Über Sinn und Zweck einer post-strukturalistischen Entgrenzung des Textbegriffs. In: *Brauchen wir einen neuen Textbegriff? Antworten auf eine Preisfrage*, hg. v. Ulla Fix, Kirsten Adamzik, Gerd Antos u. Michael Klemm, Frankfurt am Main: Lang, S. 125–141 (Forum Angewandte Linguistik; 40).
- WEMF AG für Werbemittelforschung (2005): *Wemf Auflagen Bulletin 2005*. Elektronische Ressource, <http://www.wemf.ch/> [29. 2. 2008].
- Wengeler, Martin (2003): *Topos und Diskurs: Begründung einer argumentationsanalytischen Methode und ihre Anwendung auf den Migrationsdiskurs (1960–1985)*. Tübingen: Niemeyer (Reihe Germanistische Linguistik; 244).
- (2005): Assimilation, Ansturm der Armen und die Grenze der Aufnahmefähigkeit: Bausteine einer linguistisch „integrativen“ Diskursgeschichtsschreibung. In: *Mediendiskurse. Berstandesaufnahme und Perspektiven*, hg. v. Claudia Fraas u. Michael Klemm, Frankfurt am Main: Peter Lang, S. 39–57 (Bonner Beiträge zur Medienwissenschaft; 4).
- Wittgenstein, Ludwig (1995): *Tractatus logico-philosophicus. Tagebücher 1914–1916. Philosophische Untersuchungen*. Frankfurt am Main: Suhrkamp, 2. Aufl. (Werkausgabe; 1).
- Woods, Anthony/Fletcher, Paul/Hughes, Arthur (1986): *Statistics in Language Studies*. Cambridge: Cambridge University Press.
- Ziai, Ramon/Ott, Niels (2005): Web As Corpus Toolkit. User's and Hacker's Manual. Lexical Computing Ltd., <http://www.drni.de/wac-files/doc/pre3/wac-toolkit.pdf> [29. 2. 2008].

## Nachschlagewerke

- Bußmann, Hadumod (2002): *Lexikon der Sprachwissenschaft*. Stuttgart: Kröner, 3., aktualisierte und erweiterte Aufl.
- Duden (1999): *Duden. Das große Wörterbuch der deutschen Sprache in zehn Bänden*. Mannheim, Leipzig, Wien, Zürich: Dudenverlag, 3., völlig neu bearb. und erw. Aufl.
- Duden (2002): *Redewendungen. Wörterbuch der deutschen Idiomatik*. Mannheim, Leipzig, Wien, Zürich: Dudenverlag, 2., neu bearbeitete und aktualisierte Aufl.
- Freie Enzyklopädie Wikipedia : *Liste der Hauptstädte der Welt*. [http://de.wikipedia.org/wiki/Liste\\_der\\_Hauptstaedte\\_der\\_Welt](http://de.wikipedia.org/wiki/Liste_der_Hauptstaedte_der_Welt) [29. 2. 2008].
- Grimm (1885): *Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm. Bearbeitet von Dr. Moriz Heyne*. Leipzig: Hirzel.
- Hillmann, Karl-Heinz (1994): *Wörterbuch der Soziologie*. Stuttgart: Kröner, 4. Aufl.
- Kluge (1995): *Kluge. Etymologisches Wörterbuch der deutschen Sprache*. Berlin, New York: de Gruyter, 23., erweiterte Aufl.
- Metzler (1999): *Metzler Philosophielexikon*. Stuttgart, Weimar: Metzler, 2. Aufl.
- Oxford Advanced Learner's Dictionary (1995): *Oxford Advanced Learner's Dictionary of Current English*. Oxford: Oxford University Press, 5. Aufl.
- Sprachendienst des Auswärtigen Amts (2006): Verzeichnis der Staatennamen für den amtlichen Gebrauch in der Bundesrepublik Deutschland. Stand: 18. Oktober 2006. Elektronische Ressource, Berlin, <http://www.auswaertiges-amt.de/diplo/de/Infoservice/Terminologie/Uebersicht.html> [29. 2. 2008].



## Quellentexte und Korpora

- AMTLICHES BULLETIN (o.J.): *Amtliches Bulletin der Bundesversammlung. VI/1999–IV/2003*. Elektronische Ressource <<http://www.parlament.ch/ab/frameset/d/index.htm>> [29. 2. 2008].
- Bundesamt für Kommunikation (1997): *Kultur in den Medien der SRG. Bericht des Bundesrates*. Elektronische Ressource <[http://www.bakom.ch/themen/radio\\_tv/00509/01190/01204/](http://www.bakom.ch/themen/radio_tv/00509/01190/01204/)> [29. 2. 2008].
- DEREKO IDS (o.J.): *Das Deutsche Referenzkorpus DeReKo*. Elektronische Ressource <<http://www.ids-mannheim.de/kl/projekte/korpora/>> [29. 2. 2008].
- DWDS (o.J.): *Digitales Wörterbuch der deutschen Sprache des 20. Jahrhunderts*. Elektronische Ressource <<http://www.dwds.de>> [29. 2. 2008].
- LEXISNEXIS (o.J.): *LexisNexis Deutschland GmbH*. Elektronische Ressource <<http://www.lexisnexus.de>> [29. 2. 2008].
- NZZ-KORPUS (1995–2005): *Neue Zürcher Zeitung-Korpus 1995–2005*. Elektronische Ressource [kompiliert von Noah Bubenhofer].
- POLITIKFORUM.DE-KORPUS (2002–2008): *Korpus allgemeiner politischer Diskussionen im Web*. Elektronische Ressource [kompiliert von Noah Bubenhofer und Joachim Scharloth].
- WORTSCHATZ LEIPZIG (o.J.): *Deutscher Wortschatz Universität Leipzig*. Elektronische Ressource <<http://wortschatz.uni-leipzig.de>> [29. 2. 2008].



## Stichwortverzeichnis

$\chi^2$ -Test 137

Abstrahierung 153

Alltagslogik 72

Alternativhypothese 116

Ambiguität 127

Annotation 124, 335

Anschlussfähigkeit 51

AntConc 182

Arbitrarität 114

Argumentation 153, 168, 267, 299

Argumentationsanalyse 70, 314

Argumentationsfigur 67, 85, 260

Ausdruck 63

Aussage 35, 38, 105, 309

Autopoiesis 51

Backing 69

Begriffsgeschichte 32, 36, 153

Behauptung 68

bigram 118

Britischer Kontextualismus 111

Chi-Quadrat-Test 137

Claim 68, 260

Cluster 67, 119

co-occurrences, *siehe* Kookkurrenzen

collocation, *siehe* Kollokation

Collostructions 118

Common Sense 50, 60

Common Sense-Kompetenz 63

Common Sense-Wissen 53

Concgram 118, 123

Concordancer 180

corpus-based 17, 99, 172

corpus-driven 17, 99, 123, 310, 321

corpus-driven Linguistics (CDL)  
101

corpus-illustrated 100

COSMAS II 180

Cramér's V 138

data-driven, *siehe* corpus-driven

Datum-Argument 68, 260

DDC 180

Deduktion 17, 74, 89, 102

Deutsches Referenzkorpus 175, 224

Diachronie 84

Differenzkoeffizient 144

Differenzmenge 151

Digitales Wörterbuch der Deutschen Sprache 175

Diskurs 317

Diskursanalyse

Foucault'sche 31

linguistische 32, 60

diskursive Praxis 32, 309

diskurssemantische Grundfigur 60,  
85, 313

Ensemble 25

-sachverhalt 26

-stück 25

Episteme 53

epistemisches Wissen 53

Erwartungserwartungen 58

Ethnienbezeichnung 212

Ethnokategorien 57

Exakter Test nach Fisher 142

Expressiva 331

Fachvokabular 290

Filtern (von Mehrworteinheiten)  
150

Firth, John Rupert 111

Fisher's exact Test 142

Fokuspartikel 331

Foucault, Michel 31



- Gebräuchlichkeit 116  
 Gemeinsinn 50  
 Georeferenzierung 195, 275  
 Grad der Typik 155  
 Gradpartikel 331  
 Grundgesamtheit 107, 108  
  
 Hedonismus 86  
 Hermeneutik 319, 322  
 Humboldt, Wilhelm von 46, 48, 49  
 Hypothesentest 134  
  
 idiomatische Prägung 62, 85, 309,  
     312  
 Idiomatizität 115, 336  
 Illokution 64  
 Implikation 64  
 Implikatur 85, 260  
 Implikatur, konversationelle 64  
 IMS Corpus Workbench (CWB)  
     181  
 Induktion 17, 66, 89, 102, 311  
 Interaktion 49  
 Interpretation 104, 155  
  
 Juillands D 142  
  
 Klumpen 134  
 knowledge based 100  
 Kollokation 8, 65, 82, 85, 111, 122,  
     181, 197  
 kommunikative Gattung 57, 313  
 kommunikative Muster 58  
 Kompetenz 51  
 Kompositionalität 115  
 Konkretisierung 153  
 Konstruktivismus 50  
 Kontext 27, 43, 55, 64, 66, 85, 160  
 Kontextualisierung 45, 56, 57, 63,  
     87, 89, 311  
 Kontextualisierungsprofil 160, 220,  
     232, 248, 280, 301, 303  
 Kontingenz 50, 52, 318  
 Kontingenztafel 136, 156, 157  
 Kontrastieren (von Mehrwortein-  
     heiten) 150  
 Konvention 52  
 Konventionalisierung 63  
 Kookkurrenz 8, 111, 113, 122  
  
 Kookkurrenzprofil 119, 161  
 Kooperation 49, 309  
 Korpusdefinition 37, 105  
 korpusgesteuert, *siehe* corpus-  
     driven  
 Korpuslinguistik 16, 91  
 Korrelation 134  
 Kovorkommen 113  
 Kristallisationskern 309, 319  
 Kultur 39  
 Kulturalität 44, 46  
 Kulturanalyse 84, 322  
  
 Langue 126  
 Lemma 124  
 Leserbriefe 168  
 Lexem 8  
 linguistische Diskursanalyse, *siehe*  
     Diskursanalyse  
 Log-Likelihood-Koeffizient 139  
 log-likelihood-ratio 121  
  
 Mann-Whitney-Rank-Test 139  
 Medialität 44  
 Mehrworteinheit 8, 103, 122, 181,  
     197, 310, 336  
 Menschenverstand, gesunder 50  
 Mentalität 60, 69, 313  
 Merkmalsoppositionen 61  
 Metapher 77, 85, 89, 153  
     Clausewitzsche 78  
     konzeptionelle 77, 314  
 Metaphorical Pattern Analysis 81,  
     315  
 Multi-Word Unit 122  
 Muster 8, 18–30, 92, 111  
     -ausprägung 25  
     -rezept 25  
     -sachverhalt 26  
     kommunikative 58  
     sprachliche 85  
     syntagmatische 118  
 Musterhaftigkeit 18–30, 88, 92, 320  
  
 n-Gramm 111, 118  
 negative Evidenz 91  
 Neue Zürcher Zeitung (NZZ) 189  
 Ngram Statistics Package (NSP) 197  
 Normalisierung 150

- Nullhypothese 116  
 Oberfläche 44, 89, 92, 310, 320  
 Operationalisierung 38, 310  
 paradigmatische Alternativen 56  
 paradigmatische Konstellation 163, 212  
 Parole 114, 126  
 Parsing 124  
 Part of Speech, *siehe* Wortarten  
 pattern 21  
 Phi-Koeffizient 138  
 Phraseologie 114, 336  
 Phraseologismen 114  
 Pointwise Mutual Information 142  
 Prägung 44  
 Präsupposition 64  
 Pragmatik 87, 114, 336  
 pragmatische Funktion 168, 257, 267, 302, 311  
 pragmatische Wende 15, 43  
 Precision and Recall 312  
 Redeweise 296  
 Referenzkorpus 103, 172  
 Regel 30  
 Reid, Thomas 50  
 Rekurrenz 111, 309  
 Reliabilität 319  
 Repräsentativität 33, 109  
 Rhetorik 67  
 Routineformel 65, 336  
 Routinen 63  
 Schablone 30  
 Schema 29, 63  
 Schlussregel 68  
 Schnittmenge 151  
 self organizing 100  
 Semantik 336  
 semantische Felder 61  
 semantische Matrix 320  
 semantische Oppositionen 61  
 Semiose 40  
 Signifikanz 115, 116, 121, 134, 155  
 Sorgenbarometer 250, 298  
 Sortieren (von Mehrworteinheiten) 150  
 soziales Handeln 53  
 Spinnenprofil 89  
 Sprachgebrauch 8, 15, 36, 63, 108, 320  
 Sprachgebrauchsgeschichte 84, 312  
 Sprachgebrauchskompetenz 45  
 Sprachgebrauchsmuster 37, 53, 66, 103, 296, 309, 321  
 Sprachgebrauchswissen 44, 45, 48, 311  
 Sprachgeschichte 84  
 Sprechakt 44, 65  
 Sprechen 8  
 Sprechweise 31, 32, 37, 108, 309, 317, 319, 320  
 Standardabweichung 132  
 Statistik 65, 67, 333  
     deskriptive 132  
     induktive 134  
 Stichprobe 107–109, 134, 190  
 Stil 55, 313  
 Symbolisierungsakte 84  
 Symbolisierungsmedium 84  
 syntagmatisches Muster 118, 122  
 Systemtheorie 50, 51  
 t-Test 141  
 Tagging 124, 194, 335  
 Text Encoding Initiative (TEI) 181  
 Textroutinen 63  
 Textsorte 303  
 Tiefenstruktur 44, 74, 82, 85, 89  
 Token 8  
 Topos 67, 70, 85, 314  
 Toulminschema 68, 75  
 Treffgenauigkeit und Ausbeute 312  
 trigram 118  
 Type-Token-Relation 29  
 Typik 62, 115, 320, 336  
     des Handelns 63  
     des Sprechens 63  
     Grad der 155  
     stabile 160  
     variable 160, 163  
 Typikprofil 149, 152, 199  
 Varianz 132  
 Vereinigungsmenge 151  
 Vergleichskorpus 172

- Viabilität 51
- Vorlage 23
- Wahrnehmungsmuster 51
- Warrant 68
- Web als Korpus 176
- Web as Corpus Toolkit 193
- Wetterbericht 64
- Wittgenstein, Ludwig 15, 25, 44
- Wordsmith 182
- Wort 8
- Wortarten 124
- Wortgeschichte 32
- Wortgruppen
  - diskontinuierliche 117
  - kontinuierliche 117
- von Zesen, Philipp 46
- Zufallsstichprobe 109

